FACULTAD DE ESTUDIOS ESTADÍSTICOS

MÁSTER EN MINERÍA DE DATOS E INTELIGENCIA DE NEGOCIOS

Curso 2018/2019

Trabajo de Fin de Máster

TITULO Analysis and interpretation of the determinants of parental leave use in Spain

Alumna: Cansu Ozdogan

Tutora: Alicia Pérez-Alonso

Noviembre de 2019

UNIVERSIDAD COMPLUTENSE
MADRID

Content

Figure Content

Table Content

1. Introduction

Society's growing demand for gender equality has advanced greatly over the last few years and it forms an inescapable part of the political agenda of the vast majority of western countries and democracies. According to the glossary of terms and concepts by the United Nations Children's Fund (UNICEF), gender equality means that women and men have equal rights in order to contribute and receive gains from economic, politic and social life (UNICEF, 2017). Over the past decades, significant reforms were developed by governments to provide gender equality to different life areas including family life. Noteworthy among these reforms is parental leave, which is an employee benefit for taking care of small children.

There are different types of leave that can be used by parents after the birth of a child or an adoption process. Maternity leave is the most well-known type of leave. It refers to the period of time when a mother can be absent from her work for the purpose of giving birth and taking care of her baby. This leave is usually mandatory for the mothers, but its duration varies in different countries. As the duration of the leave changes according to the country, the payment conditions change as well. Development countries offer better conditions such as getting full payment while on maternity leave. However, some developing countries just offer different percentages of the former salary of the mother.

In earlier times, maternity leave was the only right that could be used for parents to take care of their children. There was no specific leave type just for the fathers. We find different attempts to fill this gap across countries and over time. In some countries, fathers were able to use the maternity leave of the mother only if the country's policy allowed the mother to transfer this right to the father. Another option was a parental leave that could be used for both parents, but most of the time this type of leave was taken just by mothers. There was also an option of unpaid leave but given that this option could cause economic problems in the family, its take up rate ended up being really low. In the last few decades, paternity leave has been brought to some countries as another type of parental leave system, i.e. a time period when an employed father can leave his job without the risk of losing it and still getting paid for it. In the beginning, there were few countries where fathers could use this leave. Moreover, the conditions of the paternity leave (time duration, payment conditions etc.) were really limited compared to those of the maternity leave. Nowadays, paternity leave is applied in many countries. The extent of the leave is also getting better, especially in European Union (EU) countries.

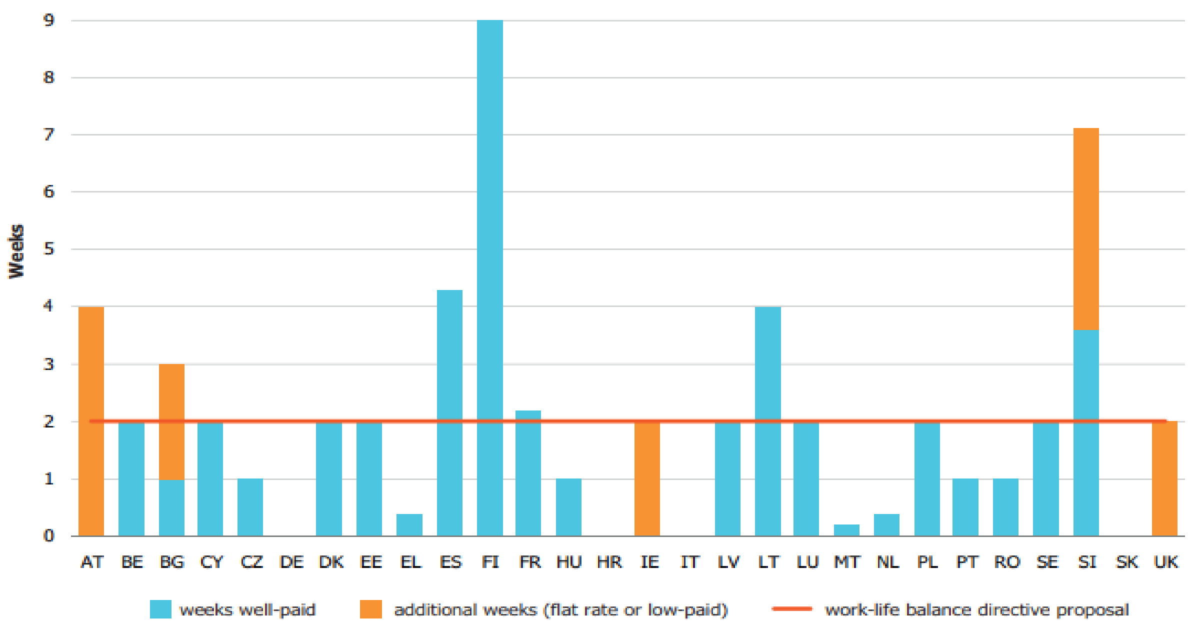
The paternity leave is an important policy that should be taken into account according to many reasons. As it was stated before, governments have applied various reforms with the purpose of bringing gender equality into their countries, and the paternity leave is a significant step that aims to bring equality inside the families. It was a known fact in previous generations that the mother was the only responsible of childcare and household activities. As the situation was like that, women had difficulties to be a part of the working life as there were not equal conditions between men and women. The labor market inequality is still affecting women in their labor life especially after the birth of a child. The term of motherhood penalty causes that women leave their employment more than men leave after a childbirth, as women with children under 12 years old get paid 11% less than women without children, when men get paid 7% more in the same case (Miani & Hoorens, 2014). Even though the gender inequality could not be removed completely, there are many policies that aim to develop gender equality in working life as well as in social life.

From the other point of view, fathers used to have a weaker role in the child-care activities because they did not have a right to leave their works for a period of time without taking the risk of losing their job and their economic source (because there was not "paid" paternity leave option). This was causing some family problems as the fathers used to have less involved relation with their children. Both mothers and fathers used to have difficulties to balance between the working life and family life, as the parental leave could not let them be a part of the responsibility of child-care in an equal way. For this reason, the paternity leave has an important role to bring a balance between mothers and fathers in the family life. Also helping more to the mothers for being more involved in the working life as their responsibility could be distributed with the fathers.

According to a recent research of the World Bank, among 187 countries that were analyzed, those countries that have some type of parental leave that parents can take after having a child increased from 77 in 2009 to 107 in 2018, with an increase of 39%. Most of them have paternity leave, and the rest has a parental leave that allows both parents to use the shared right (Borraz, Sánchez & Ordaz, 2019, as cited in World Bank, 2018).

In the case of European countries, duration and payment regulations differs depending on the leave policy of those countries. According to a research that was realized by the European Platform for Investing in Children (EPIC) in 2018, there are 17 European countries that reached to the 2-week paternity leave. However, there are just 13 countries which offer a compensable payment condition during the leave. Figure 1 shows the length of the paternity leave in weeks in the EU countries.

Figure 1 Paternity leave on Countries Source: (EPIC,2018)



As it can be seen in Figure 1, the country which has the longest length of paternity leave is Finland with 9 weeks of leave. The blue color represents the well-paid weeks, and it indicates a payment that is at least 66% of the former salary of the father. After Finland, the second country that has the longest paternity leave is Slovenia. However, only the first four weeks are well-paid and the rest of the weeks are low paid. In the case of Spain, it can be seen that the country has a longer leave duration compared to most of the countries in the figure. Fathers from Spain had four weeks of paternity leave in 2018, and during the leave the fathers were well-paid. The case of Spain and the reforms will be analyzed deeper.

All tables and figures are own elaborated unless stated otherwise.

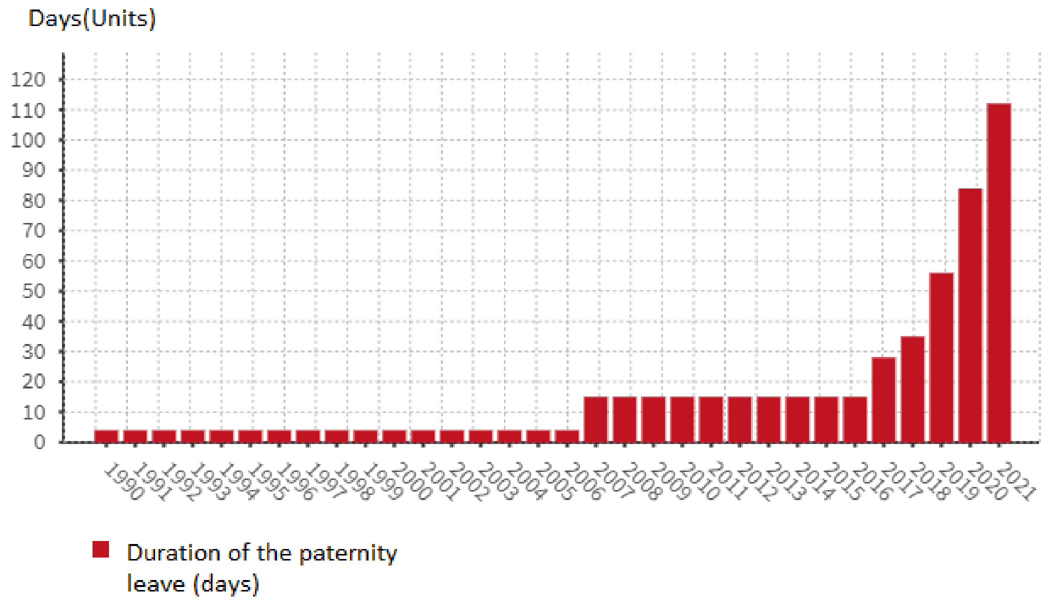
2. Paternity Leave Reforms in Spain

In Spain, there was not a paternity leave right until the 2000's. This right was first introduced in March 2007 and gave the fathers 13 days of permission. The leave was uninterrupted, and it could be taken full-time or part-time. During the leave, fathers were entitled to the entire amount of their salary. In addition to the 13 days given by the government, employed workers had 2 additional days paid by the company. Thus, there were 15 days of leave in total (Escot, Fernandez-Conejo & Poza, 2013).

Figure 2 Paternity leave reforms in Spain

After the first reform was made, there was a possibility to extend it to one month in 2011. But as the economic crisis negatively affected the economic and labor conditions of the country, it could not be extended in the year that it was planned. The second reform entered into force in 2017, after 10 years of the first reform. In January 2017, 4 weeks of paternity leave were brought for wage earner men who have become fathers after the first day of 2017. As it can be observed in Figure 2, there was a long time between the first reform and the second reform, but the third reform entered into force a short time after the second reform. After one year and a half, in July 2018, the third reform was implemented, and the paternity leave increased its duration in one week, from 4 to 5 weeks. The leave duration did not change a lot until the latest reform, which is currently in force. The 1st of April of 2019, the fourth and last reform was carried out, and the leave duration was extended from 5 to 8 weeks of permission. Figure 3 shows the evolution of the duration of the paternity leave over the years. As it can be seen, since 2017 there are modifications in the length of the leave, and also new extensions are expected for 2020 and 2021.

Figure 3 Duration of the paternity leave in Spain Source: (OCDE, Sanidad, 2019)



Requirements of the paternity leave are related with the Social Security Institution of Spain. Two fundamental requirements are to be registered in the Social Security system and the father should have a contribution period of 180 days within 7 years prior to the start date of the permit, or 360 days throughout his working life (Seguridad Social, n.d.).

3. Literature Review

In 2018, the WORLD Policy analysis center presented a research about paid parental leave in OECD countries (WORLD, 2018). It states that 33 of 34 OECD countries provide paid leave, except United States. The research remarks some advantages of the paid parental permissions on supporting women´s economic opportunities and gender equality at home when these leaves are available for both genders. WORLD (2018) carried out a cross-national analysis that shows that when the duration is longer, and the compensation level of these permissions is getting higher, it helps the participation level of women in labor force. Moreover, it was found that a longer parental permission helps to reduce the employment gap between mothers and fathers. Akgunduz & Platenga (2011) performed a study that also states that a parental leave with strong sides of job protecting aspect brings mothers into labor force and increases the participation. This study is based on the methodology of difference-in-differences.

The duration of the leave is important but also the compensation rate affects highly the decision of taking a leave. Zhelkazkova (2013) highlights the wage aspect with the analysis of working parents in Luxembourg. The analysis shows that fathers who have higher income compared to their partners are less likely to use a parental permission. According to the same aspect, a study that was done in Germany about dual-earner couples shows that the fathers who have lower income than his partners are more likely to take a parental leave, compared to the fathers who have a similar level of income with their partners (Reich, 2011).

Meil et al. (2010) showed that, in recent years, parental leaves are extended, and that motivates fathers to use more these leaves to be more responsible in childcare activities and to bring gender equality in family life. Despite the extended leaves, there is a large difference in the take-up rate of parental leaves between males and females. When the fact is analyzed deeper to figure out the reasons why there is a large difference on the use of parental leaves between both genders, it was found there are some obstacles that affect the take-up rate of the fathers. Belle (2016) states that there are conditions that might change the willingness of the fathers to use these permissions, such as level of compensation of the leaves, eligibility (only wage earner employees can take leave), employers' support, company culture, society norms, etc. Mail et al. (2010) expanded the list by indicating that the take-up rate differentiates with the type of contract, education level and the conditions of job instability. Especially, self-employed parents or parents with temporary working contracts have lower take-up rates and lower salary conditions.

Escot et al. (2013) investigated the effect of a 2-week paternity leave and analyzed some conditions of the fathers in order to see how they impacted the take-up rate of the permission between 2005 and 2009. They found that there is a statistical evidence, which proves that fathers use more the parental leaves after 2007 when the 2-week leave entered into force. The investigation pointed out that fathers are more likely to take the parental permission if they are working in public sector, if the partner has a job and if they have stability in their working life. During the study, other conditions were investigated, such as demographic, social and professional factors of the individuals. Since the last statistical analysis was carried out about the 2-week permission, the paternity leave was extended three times. As there is not a current analysis that investigates the new changes of the paternity leave, we are going to carry out that investigation by extending the time scope of the analysis from 2005 to 2018, increasing the comprehension of the variables and focusing on the changes of the policy.

4. Project Justification

As it was mentioned before, lately there are many reforms that aim to motivate men to be more active in family life and childcare activities. These reforms also affect positively that women can be more integrated in labor life after the birth of child, as mothers can get back to work without facing some common difficulties, such as being the only responsible of the children or losing the status of wage earner member of the family. Although there are many regulations about the paternity leave worldwide, the gap between the use of maternity leave and the use of paternity leave is still quite big.

First of all, we want to find out if the gender gap on the take-up rate of the parental leave is changing since the first reform of the paternity leave was implemented in Spain. In order to do that, we want to see if there is a positive tendency among fathers to use the parental leave after the paternity reforms. Moreover, we would like to analyze if the gap is reducing while the reforms are modifying and the length of the paternity leave is extending in time.

Another concern of the project is to figure out in which conditions fathers are more willing to use the paternity leave. The life conditions of parents will be analyzed to see which factors are influencing the decision of taking the parental leave on fathers and mothers. The target group is going to be fathers. The mothers will be used as a control group for being able to compare if the analyzed conditions are just affecting the fathers or both genders. During the project, it will be used a survey called The Economically Active Population Survey, which enables us to access various variables to analyze the conditions of the fathers, such as demographic data, working conditions, and educational and professional life's factors. As the survey is conducted to all the members of the household, we are able to analyze the conditions of the mothers as well as the conditions of the fathers.

4.1 Economically Active Population Survey

The Economically Active Population Survey (EAPS) is made by Spanish Statistical Office (INE, in its Spanish acronym) and it was first published in 1964. It is a continuous survey on a quarterly time period, and it contains data on labor market in Spain. The target population of the survey is the population living in family households. In the case of a foreigner living in the household that is interviewed, the foreigner is included in the survey if he/she lives in Spain more than 1 year. The method of carrying out the survey is personal or phone interviews. Survey questions are asked to get information about a specific period of time which is referring one week before the interview. The survey is done in the all-Spanish territory (INE, n.d.).

On the website of INE, data description is provided and the main characteristics of the survey are shown below as they are presented in the official page.

- Employment according to demographic variables (sex, age, nationality, marital status, education), professional status, underemployment, working hours, type of working day, type of contract, having more than one job, etc.
- Unemployment according to demographic variables, characteristics of the previous job, search methods, search duration, etc.
- Besides the aforementioned information, there is an annual publication of information regarding 'sub-sample variables': studies sector, specific work conditions (shifts, weekend jobs, persons who work in the establishment, supervision responsibilities etc.) and the characteristics of the last job of the unemployed persons with previous professional experience (INE, n.d.).

5. Data Preparation

As there is no direct official survey about parental permission in Spain, we used an official survey that already exist, the EAPS, to build a database that can be useful for analyzing the characteristics of parents and the use of maternal and paternity permissions. Thus, a data preparation process was necessary for extracting the meaningful part of the data before making any statistical analysis about the parental permission. The microdata of the survey can be found in the Microdata Section of the INE website from 2005 until now. As it was explained before, this survey aims to collect data from families about the labor force, employment and unemployment situations as well as the demographic data of each family member in a household. In quarterly data, there are approximately 65,000 families and it refers to 180,000 individuals/family members.

To illustrate the source code, we used for the data preparation process we choose the first quarter of 2018 as reference. The first quarter of 2018 had 167536 observations and 93 variables before starting the data preparation process. The algorithm described below was applied to each quarter and a depurated sample was obtained. At last step we merged the resulted depurated samples from 2005 to 2018. The data preparation process was done in RStudio. The whole Data Preparation process with all the steps of the algorithm that are listed and explained below can be found in Appendix A.

STEP 1- Identifying the children less than 5 years old

The first step was identifying households where there are children who are less than 5 years old. In the survey design, the age variable has categories for every 4 years (0-4 years old, 5-9 years old, etc.). Thus, we could only identify children who are less than 5 years old. We could not identify children who are less than 1 years old or who are just born to actually create a more specialized sample with the households that have the right to use the parental (maternity and paternity) permission. In this step, another filter was also applied to choose the households where the children have both parents registered (living) in the house. The reason of the second filter is to select and stay with the houses that are both parents live in which aims to make a better comparison later between the mothers and the fathers in the statistical analysis. In order to do that, the total observations were taken from the one quarter microdata and it was created a filtered data with only the children who is less than 5 years old and who has the mother and the father in the house. In the end of the filters, for the first quarter of 2018, we could detect 6388 children who fulfill the required conditions. The process was done with the filter function of *dplyr* (a data manipulation package) in R.

STEP 2- Selection of the households with all the members from the previous step

From the filtered data just with children, we extracted a variable called NVIVI, which is a variable for identifying each household. We have a list of households where children who are between 0 and 4 years old live with both parents. After extracting the identical (id) numbers of the houses, we realized that if there were more than 1 child in the house who fulfilled the conditions (0-4 years old, have both parents at home), the register of the house would be duplicated or triplicated. Thus, we decided to keep just one register of each house, even if there was more than one child. As the final data set will just contain the observations from the parents and not the children, we would not lose any information by deleting a child register from the list of the households, because we could keep the data of the parents from those houses. After applying this filter, the number of observations decreased from 6388 to 5462. Thus, we could confirm that there were households that had more than 1 child who lived in the same house with another child, and who had the conditions that we were looking for. To sum up, the aim of this step was to create a data table that contains the identification number of the houses where there were parents who have children (between 0 and 4 years old) in common. In this way, we will be able to compare the rate of use of parental permission among fathers and the rate of use of parental permission among mothers in the end of the creation of the dataset .We then used the *inner join* function of *dplyr* to merge the original microdata with this data table. We ended up with a data table that contains all the members of the households where there are children between 0-4 years who live with both parents.

STEP 3- Creating a new variable for identifying the number of children of each parent

We created a new variable for the fathers to identify how many children (less than 5 years old) they have. For doing that, we used an original variable from the microdata called NPADRE, which shows the identical number of fathers in the households. In the survey design, the interviewer assigns an id number to each member of the household. For instance, if there are 5 people living in the house it means that the members of the house have id numbers from 1 to 5. That identical number is called NPERS and it is a variable that is going to be used in the following steps. For instance, if a person has the id number 5, and if this person is the father of a child from the same household, the child will have the id number of his father (NPADRE) with the same number that is 5. It was important to use the variable NPADRE because there could be more than one father in a household, and we need to identify each father.

The data table which has the register of all the children was taken for creating a new table from that. We grouped the houses with the household identical number (NVIVI) and the identical number of the fathers (NPADRE). After that, we created a new variable that gives the total number of children of each father who has a different father identical number in the household. Table 1 displays the summary of the new data table with the new variable.

Table 1 Number of fathers vs. Number of children they have

	Number of children		
	1	2	3
N of fathers	4596	836	40

For the first quarter of 2018, Table 1 shows that there are 4596 fathers who have 1 child between 0-4 years old who lives with his/her parents. And with the same conditions, 836 fathers have 2 children and 40 fathers have 3 children.

We used the *left join* function of *dplyr* for joining the data table of households (all the members included) with the new created table that only contains the fathers. For joining the tables with the left join function, we considered the id number of the households (NVIVI), and whether the id number of the household member (NPERS) matches with the id number of the father (NPADRE).

We replicate the same process for the mothers using the identical number of mothers in the household. The function *left join* was used again for joining the previous data table with the new created table that only contains the mothers. We ended up with a new table that has the registers of the household members and the number of children of the fathers and of the mothers.

Using this last table, we created a new variable with the *mutate* function of R. It is a common variable for both parents that shows the total number of children of each person. The two previous variables of the fathers and the mothers that contained the same information as this common variable were deleted from the data table.

STEP 4 – Researching the extraordinary cases in the data table and identifying the key variables for detecting them

After checking the last data table that was created in Step 3, we realized that the database contained a few extraordinary cases. For instance, in some households, there were parents who had a common child but one of the parents had another partner in the same household, according to the survey registrations. For avoiding extraordinary situations and some possible mistakes that could be done while the survey was registered in the microdata, we made some extra checks to ensure that the final data set only contains parents who have children in common, and these parents are registered as partners.

To this end we created new variables for detecting extraordinary cases with two key points. First one has to detect if there is a person who has a child but who has not a registered partner in the household. Second one has to detect if the data table has the register of a person who does not have a child. As the final data set should only contain the registers of the fathers and the mothers, there should not be any register that has 0 as its value from the variable that records the total number of children.

First variable has been created for the "no partner cases". The variable was created as a binary variable, which is called NOPAR, to identify if there is an extraordinary situation in the household. It takes value 1 if such a situation is identified and zero otherwise. We next created another variable to record how many people in the household has a 1 in the variable NOPAR. For creating this variable, we first grouped the data table by the variable NVIVI. We used the function *summary* for summing the number of NOPAR=1 that were registered for each household. We then filtered the households that have the value 0 in the summary of NOPAR. In other words, we kept the households that do not have "no partner" situation. After filtering the data, we deleted the variables NOPAR and its summary because they are not going to be used in the analysis.

For identifying the children (between 0-4 years old) who do not live with their parents and removing them from the database, we have created a new variable that identifies an individual who has a partner in the house but do not have a child. For identifying those partners without children, it has been created the new variable which was used for removing the explained cases from the final dataset. After creating the variable, the filtering process was applied for removing the members who do not have children (0-4 years old). In the end of the filter process, only the individuals who have a partner with whom they have a child who lives with them in the household remained.

STEP 5- Making a loop in R for merging all filtered microdata

We checked the source code of the data preparation with the reference quarter. After confirming that it was working, we programmed a loop r to apply the source code to each quarterly data. When we had all the filtered quarterly data from 2005 until 2018 in separated R files, we merged them into one R file. We created a data set with 632362 observations at the end of the process.

STEP 6- Filtering the data for maintaining only wage earner parents

We restrict the statistical analysis to wage earners. In the survey, there is a variable that gives information regarding the working situation of the person. The variable has 7 categories, which are defined as businessmen with employees; independent worker; a cooperative member; worker in a family business; waged earners in public sector; waged earners in private sector, and others. Due to the design of survey, if the interviewed person is neither a wage earner from public sector nor a wage earner form private sectors, this person is leaded to skip some questions from the job section, such as conditions about the working contract, working hours, contract type, etc., that we want to analyze. As these people are leaded to skip these relevant questions, it causes that the corresponding variables have missing values from the interviewed people who are not wage earners.

Moreover, we preferred to have a more homogeneous dataset that only includes workers who have jobs based on working contracts. The filter was applied according to the variable that indicates what kind of working situation the person had during the reference week. After applying this filter, the number observations decreased from 632362 to 381073.

STEP 7- Creating a dependent variable for the usage of permissions

The EAPS does not have any variable that we can use as our target variable for the parental permission. But we found two target questions in the survey that were useful for creating our target variable.

According to the design of the survey, there are many questions that are designed as chain questions. It means that depending on the responses of the person interviewed, he/she is derived to answer different questions or even different sections of the survey. As an example, if the interviewed person does not have a job, the interviewer skips the section about the working conditions and moves to another section about the job research conditions. In the opposite case, if the interviewed person has a job, he/she will answer the section of working conditions but not the section of job research conditions.

We next stated the first question form the section Relationship with Economic Activity of the survey:

"The reference week, Monday through Sunday, did you work even if it was just an hour?"

If the answer is no, the interviewed person is asked to answer the following question:

"The reference week, did you help in the company, business or operation of a family with whom you live without being paid for it?"

If the answer is no again, then the next question is asked:

"Did you have a job or business during the reference week, even if you didn't work in it?

If this time the answer is yes, we reach the first target question for being able to pursue permission cases.

"What is the main reason you did not work in that job?"

For answering to this question, there are fourteen options that can be chosen. Two of them are about parental permissions, which are stated as birth permit for a child and unpaid leave for a birth of child. As it was already mentioned, paternity and maternity permissions are paid permissions in Spain, but we are going to include the answer of "Unpaid leave for a birth of child" as a condition of the target variable. Because, in the end, we would like to see the gap on the usage of permissions that were taken with the aim of taking care of newborn babies among all the mothers and fathers of the database. For this reason, it is not a problem that we add both answers as conditions of the target variable.

The second target question was found in the section of characteristics features of the job. To reach this section, people are asked if they have a job in the reference week and if they have; the interviewer is allowed to ask the following questions about the people's job.

Before reaching the second target question, the interviewed people are asked:

"How many hours per week do you usually work in this job?"

The person is answering the question depending on his/her weekly working hours. After that the person is asked the second question that is stated in the survey as it can be seen below:

"In the reference week, how many hours did you work in this job?"

After answering the second question, if there is a difference between the regular weekly working hours and the hours that he/she has been working during the reference week, the interviewer is asking the question that is our second target question for creating the permission variable. The question is stated in the survey as:

"What is the main reason you worked a different number of hours than usual?"

There are 15 options for answering the second target question. One of them is related with the permission and it is presented as "Birth permit for a child". As the answer is directly related with the cases we are looking for, we used this question as one of the target questions for preparing our target variable. Thus, the answer of birth permit is used as one of the conditions under which our target variable has the event (permission=1).

STEP 8- Creating a variable for changes of the paternity leave policy

The database has microdata from 2005 until the last quarter of 2018. During this time period, some changes in the length of the paternity leave occurred. From 2005 to 2007 there was not a paternity permission, but since 2007 the paternity leave is in force. Therefore, we have a database that includes observations when there was no paternity permission but also observations when there were 2, 4 or 5 weeks of paternity permission. For data analysis, it will be important to have a variable that signals the change of the policy, because we would like to study if a raise of the paternity permission would positively affect the fathers to use the permission.

For the evaluation of the permission, we created a new variable named CAMBIOPOLI with four categories: NOPERMISSION, 2WEEKS, 4WEEKS and 5WEEKS. As we have quarterly data, we distributed the quarterly data according to the dates when a paternity leave reform entered into force.

Figure 4 The change of the paternity leave

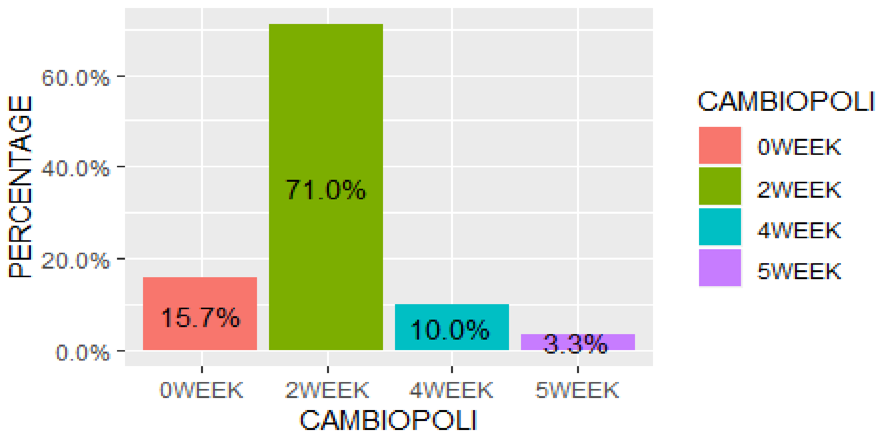


Figure 4 shows that the paternity policy that has the highest percentage of observations (71% of the database) is the first reform, when the length of the leave was 2 weeks. The smallest percentage of data corresponds to the last reform, when the government extended the paternity leave to 5 weeks.

As our data finishes in the last quarter of 2018, we only observe a period of three quarters for the 5-week paternity policy. This corresponds to the 3.3% of the observations. We decided to remove these observations due to the lack of representativeness. Thus, we used another filter to remove the observations from the last reform (the three last quarterly micro data). The database decreased from 381073 to 368455 observations.

To sum up, we keep 2 changes of the paternity leave policy. The first change is from no paternity permission to 2-week paternity permission, and the second change is from 2-week to 4-week paternity permission. We created two reference variables for observing the effect of each policy change separately. FIRSTPERMISSION is a binary variable which has its value “0” when there is no permission, and which has its value “1” since the first reform (2 weeks) entered into force.

Similarly, we created the binary variable SECOND PERMISSION. The observations have the “0” value until the second reform (4 weeks), and after that they have the “1” value.

Below we displayed a summary of both variables.

```
> table(DATABASE$FIRSTPERMISSION,DATABASE$CAMBIOPOLI)

     0WEEK   2WEEK   4WEEK
  0  59534       0       0
  1      0  270735   38186
```

```
> table(DATABASE$SECONDPERMISSION,DATABASE$CAMBIOPOLI)

     0WEEK   2WEEK   4WEEK
  0  59534  270735       0
  1      0       0   38186
```

STEP 9-Adding new columns (the partners' variables) in the same row of each observation

For each register of the dataset, we add new columns to record information relative to the partner. Thus, we can observe information relative to both parents in the same line. We first add a new column for identifying the id number of the partner in the household. After creating the identical variable of the partner, we programmed a function that takes the selected variables from the observation line of the partner according to his/her identical number in the household. The relevant features of the partner that were selected are explained in Section 7.

6. Data Sample

The final database has 368455 observations. Although it is more preferable to use all of the observations for carrying out the statistical analysis, it was not possible due to computer capacity constraints related to SAS Miner, as the size of the data is quiet big. Therefore, we decided to use a data sample instead.

We were concerned foremost with not losing any relevant and representative information from the database while constructing the data sample. First, we selected a random sample, which gives to every observation an equal opportunity of being chosen. A random sample with 10000 observations was taken. However, the random sample did not select observations from some quarterly data where there were not many observations with (permission=1). We decided to change the sampling method because we do not want to lose observations that have (permission=1).

Table 2 shows how the target variable (the variable permission) is distributed in the database.

Table 2 The take up rate of permission in the total data

PERMISSION	N of observations	Percentage
1	355114	0.96379205
0	13441	0.03620795

The distribution of the target variable is really unbalanced, as only 3% of the total data have (permission=1). As there is a huge imbalance in the categories of the dependent variable, we decided to use under sampling. Under this sampling scheme, we maintain all the observations from the category that the minority has, and we take the same amount of observations from the category of the majority. To perform that, we use the under-sampling function from the package *caret* in RStudio. We finally created a data sample that has 26.682 observations, which is double the number of observations from the minority category of the dependent variable.

We next compare some relevant aspects of the database and the data sample. On the one hand, we analyze if there was a change in the use of the parental permission by the fathers over time. Between 2005 and 2007 there was no paternity permission, but it was possible to use the transferable part of the maternity permission or the unpaid birth permissions that we included in our database. The percentage of use was calculated by counting the number of fathers who are using a permission and dividing it by the total number of fathers in the database.

Figure 5 Use of the permission by fathers in the database

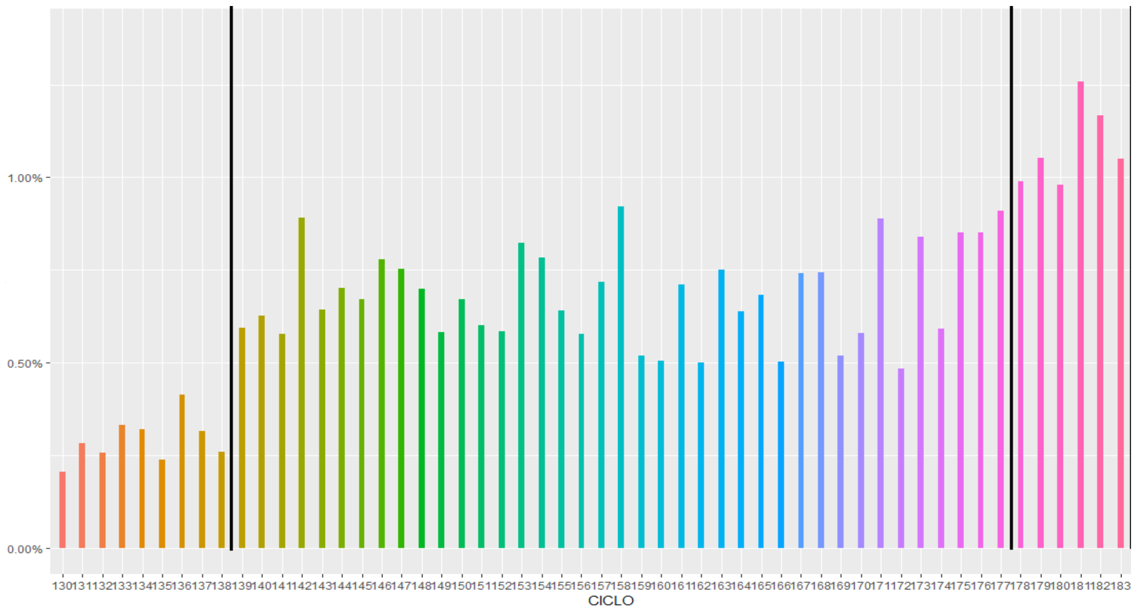
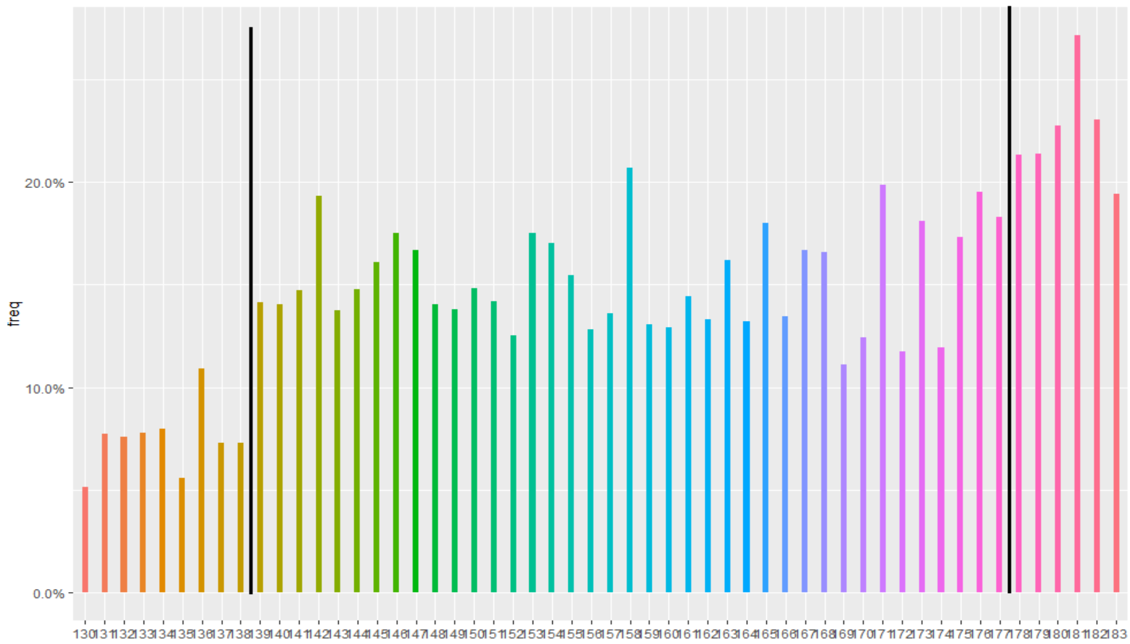


Figure 5 was divided into three parts with black vertical lines for emphasizing periods of reform change. In the vertical axis, we have the frequency of use of the permission and, in the horizontal axis, we have time, measured quarterly. From the period of "No permission" to the period of "2-week permission", there is a clear raise in the rates of the parental use among the fathers. From the period of "2-week permission" to the period of "4-week permission", we also observe that there is a raise of the frequency rate.

Thus, fathers respond positively to reform changes, as there is a clear raise in the take up rate of the permission. This fact has to be replicated in the data sample. We pay attention to that detail because it is one of the critical points of our project. We would like to analyze whether the fathers are responding to the paternity reforms, and also in which conditions they are more willing to use the leave. Hence, it is important to check if we can maintain in the data sample the rise that we observed in the database.

For comparing the previous graph with the new data sample, we have created the following graph (Figure 6) in the same way that we used before for the first graph, but this time with the new dataset of 26886 observations. As it can be observed, it seems Figure 6 shows that the data sample is still representative for replicating the raise in the use of the parental permissions among the fathers. The rate of the use of the permissions are changed as we reduced the number of observations and created a sample with under sampling method. However, the raise of using the permissions between the paternity reforms still can be seen clearly.

Figure 6 Use of the permission by fathers in the data sample



On the other hand, we analyze the gender gap on the take-up rate of the permission. We had some limitations while calculating the take-up rate because we only had information about the parents who were using any parental permission during the reference week of the survey. But it was not possible to identify the parents who were not using any permission even though they had babies, and thus they had right to the parental permission. Another limitation relates to the time period of the survey. If a parent took the permission one month before being interviewed, we would not be able to detect him/her in the survey. Parents will answer about their situations in the reference week, but not in the previous month. Because of these limitations, we calculate the take-up rate by dividing the number of fathers using permission by the total number of fathers of the database. The take-up rate of the permission for the mothers was calculated in the same way. We can only compare the take-up rate between genders as calculated. Different results on the usage of the parental permission might shed light on behavioral differences between genders.

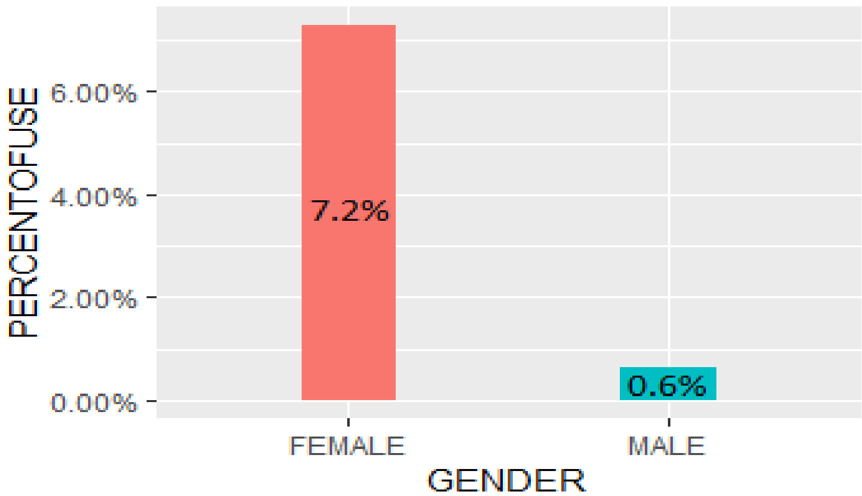
Table 3 shows the number of observations on the use of the parental permission in the database by gender.

Table 3 Number of observations in the database

GENDER	PERMISSION	FREQUENCY
FEMALE	NO	147148
FEMALE	YES	11968
MALE	NO	207966
MALE	YES	1373

After that, we have created a bar plot for comparing the take-up rates between the mothers and the fathers. The plot can be seen below.

Figure 7 Gender gap on the usage of the permission in the database



According to Figure 7, in the database, the percentage of usage of the parental permission by fathers is 0.6%, and by mothers is 7%. Thus, mothers have a take-up rate that is almost ten times bigger than the take-up rate of the fathers. We can infer from this graph that mothers are more likely than fathers to take the parental permission.

We next reproduce both figures for the data sample.

Table 4 Number of observations in the data sample

GENDER	PERMISSION	FREQUENCY
FEMALE	NO	5546
FEMALE	YES	11968
MALE	NO	7795
MALE	YES	1373

We observe from Table 4 that, for the parents who have permission in the data sample, we have the same frequency numbers because the sample was created with the under sample method. But the frequencies of the parents who do not have permission were reduced approximately by one third for each gender.

Figure 8 Gender gap on the usage of the permission in the data sample

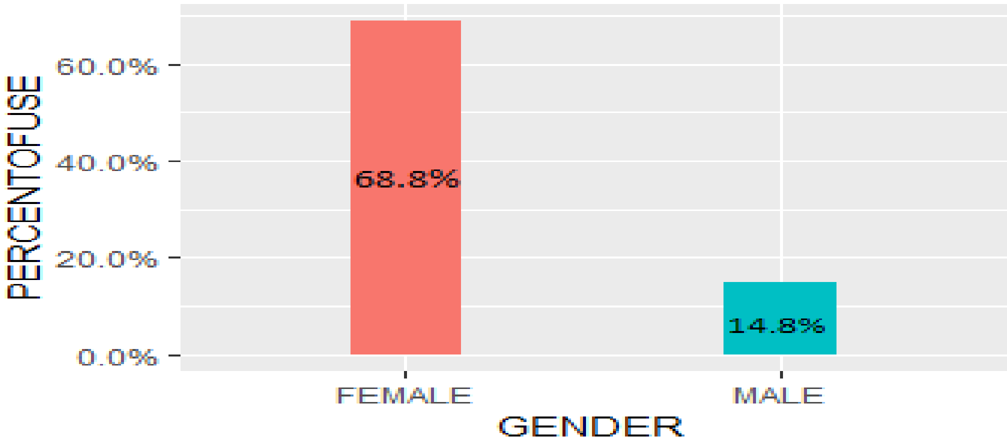


Figure 8 shows that the take-up rates changed compared to the rates of the database. However, the gap between the mothers and the fathers on the use of the parental permission is still maintained. After a childbirth, mothers utilize the permission much more than fathers.

7. Description and Exploration of Selected Variables

The INE survey has originally 93 variables. The original variables of the survey were categorized into different groups: variables for data controlling (the code number of a household, the order number of a family member in a household, indicator for adult or minor members of a household); variables of demographic data; variables of education and the level of study; working situation in the reference week; variables about the principal work; variables about the second work in case that a member has it; variables about job search in case of a member is looking for a job; variables about professional experience, and some variables about the situation of the interviewed person in the last year.

We added 11 variables during data preparation, 7 of them were partner's variables. The reason for adding these variables was to create a dataset where we could see the relevant information of the partner in the same register (in the same row) of the reference person. The other four variables were the number of children that the person have, policy change of the paternity leave, first reform identification and second reform identification. Thus, we ended up with 104 variables in our dataset.

Before using any statistical model, we did a pre-selection of the INE variables from our sample data. This selection was required because some of these variables did not have any connection to our research question as the survey contains information from different areas. Therefore, we chose the variables that might impact the use of paternity or maternity permissions. Below, there are descriptions of these potential variables together with the partner's variables whose description was not included in section 5.

7.1 Weighting Factor

Before starting to describe the selected variables, we are going to mention about the key variable that we have used for the descriptive analysis that can be seen below in each section of the selected variables. The survey of INE has the variable weighting/raising factor which is quite important for bringing the survey data to a representative level for representing the Spanish population. According to Pacific Community (2019), weighting factor is used to multiply unit of analysis to representative level. The factor has a big importance especially in the data sets that are prepared from surveys and are representing a small sample of the total population. For that matter, it was essential to use the weighting factor while creating the frequency distribution graphs for each variable as they are presented below in the section of each selected variable.

In the documentation of the survey database on the website of INE, it has been stated the method that is used for creating the weighting factor. According to INE, once the interviews have been completed, the first phase is reviewing and debugging the collected data. After the first phase, the identification variable of households (NVIVI as it was mentioned before) and certain demographic variables such as the age variable have been used for the correct assignment of elevation factors (INE, n.d.).

In order to carry out a correct descriptive analysis, we used the weighting factor in all the graphs we created for the frequency distribution of the variables on their different categories. The R code used for making the descriptive analysis is displayed in Appendix B.

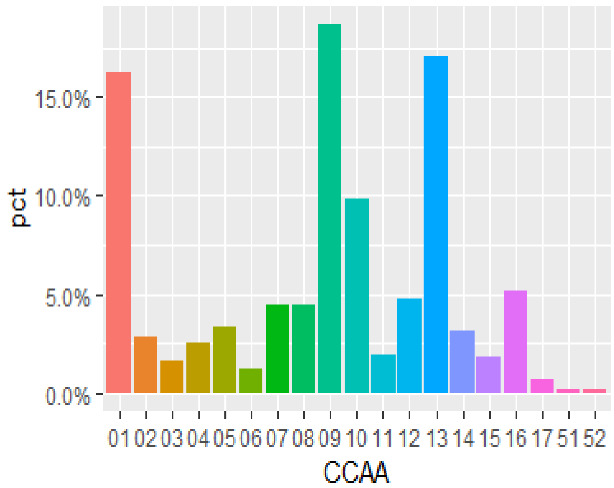
- *CYCLE (CICLO in the microdata)*

This variable is a year-quarter identifier. The values start at 130, which represents the first quarter data of 2005, and continue until 186, which represents the first quarter of 2019, the last quarter available. It is a trend variable that aims at capturing any change in the use of the parental permissions.

- *AUTONOMOUS COMMUNITIES (CCAA in the microdata)*

This variable represents the Autonomous Communities of Spain. In previous analyses, it was already mentioned that there were some encouraging regions where employees had better conditions while using parental permissions. Thus, this variable could affect the take-up rate of parental permissions. Figure 9 shows the distribution of our data sample among all Autonomous Communities in Spain. The communities that have more observations are Andalusia, Catalonia and the Community of Madrid with more than 15% of the data.

Figure 9 Autonomous Communities

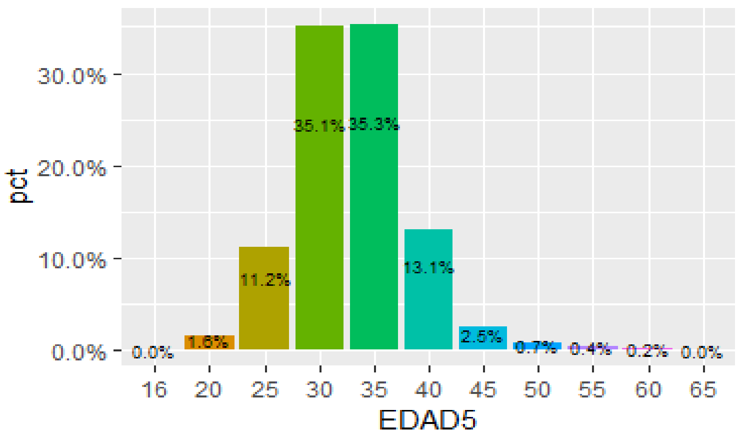


01	Andalucía
02	Aragón
03	Asturias, Principado de
04	Balears, Illes
05	Canarias
06	Cantabria
07	Castilla y León
08	Castilla - La Mancha
09	Cataluña
10	Comunitat Valenciana
11	Extremadura
12	Galicia
13	Madrid, Comunidad de
14	Murcia, Región de
15	Navarra, Comunidad Foral de
16	País Vasco
17	Rioja, La
18	Ceuta
19	Melilla

- *AGE (EDAD5 in the microdata)*

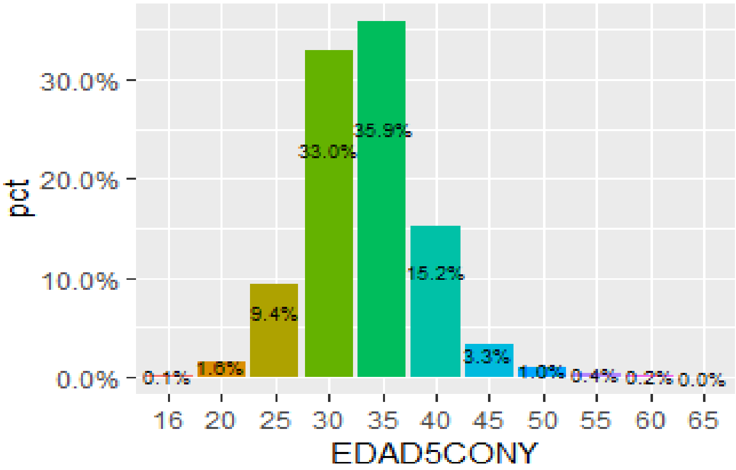
This variable is arranged by five-year age groups. Age could be important in the analysis as the use of the parental permission could change between young and old parents. For instance, we would expect the age factor to have a negative effect in the decision of using the paternity permission, as older fathers have, in general, better job positions, and thus they might have a higher risk of losing a promotion and/or they might have more responsibilities, which makes more difficult for them to leave the position. Figure 10 shows a bar plot of age variable. Most of the parents are between the age ranges of 30-34 to 35-39 years old.

Figure 10 Age Variable



The age variable has been created for the partners as well for adding to each observation in the data sample as it was explained before. In other words, we extracted the age variable from the row of the partner in the data in order to put it in to row of the person that we are analyzing. We also consider the age of the partner as a potential explanatory variable given that it might influence the decision of taking the parental permissions. Partner's age is named EDAD5CONY and its distribution is showed in Figure 11. EDAD5CONY distribution mimics the shape of EDAD5.

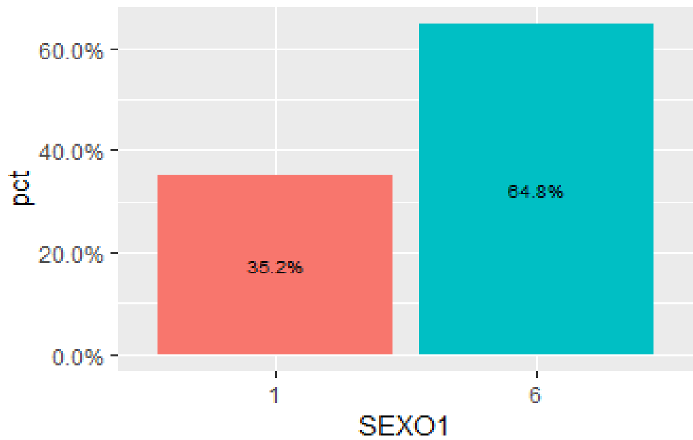
Figure 11 Age Variable for the Partners



- *GENDER (SEXO1 in the microdata)*

The gender variable is a binary variable that takes value 1 for males and value 6 for females. After the data preparation, the database only contains mothers and fathers who have children less than 5 years old in the reference week. The data sample was created with sets of parents who share a common child. It is important to keep the mothers in our data sample because we are going to use them as the control group while investigating the characteristics of the fathers who use the parental permissions. Figure 12 shows that mothers (control group) represent 64.8% of the data and fathers the remaining 35.2%.

Figure 12 Gender Variable



- *CIVIL STATUS (ECIV1 in the microdata)*

By including this variable, we want to check whether civil status affects a parent's decision on taking the corresponding parental permissions even though he/she has a partner in the household. The variable has 4 categories: single with value 1; married with value 2; widow with value 3 and separated or divorced with value 4.

Figure 13 Civil Status Variable

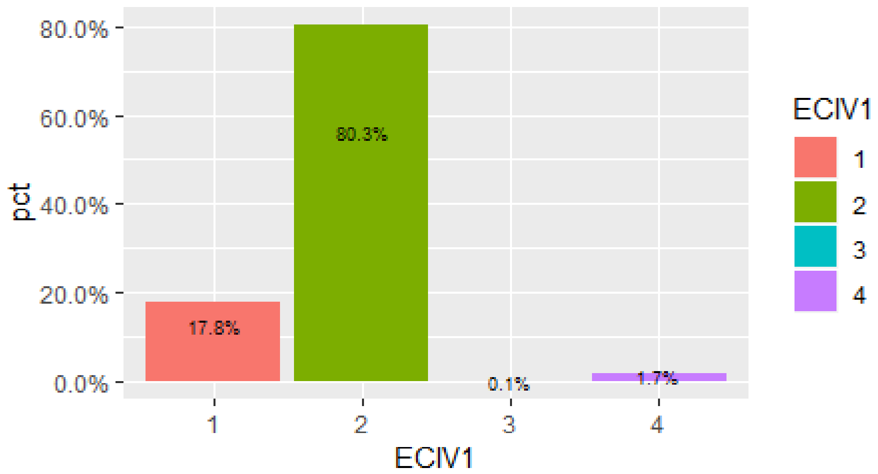


Figure 13 shows that most of the parents who are represented in the data sample are married with the percentage of 80.3% among all the observations.

- *NATIONALITY (NAC1 in the microdata)*

The nationality of the parents was included as an immigrant situation might cause job instability and it could impact negatively on the individual decision of using the parental permission. The variable has 3 categories; Spanish nationality has the value 1, Double nationality has the value 2, Foreigner has the value 3.

Figure 14 Nationality Variable

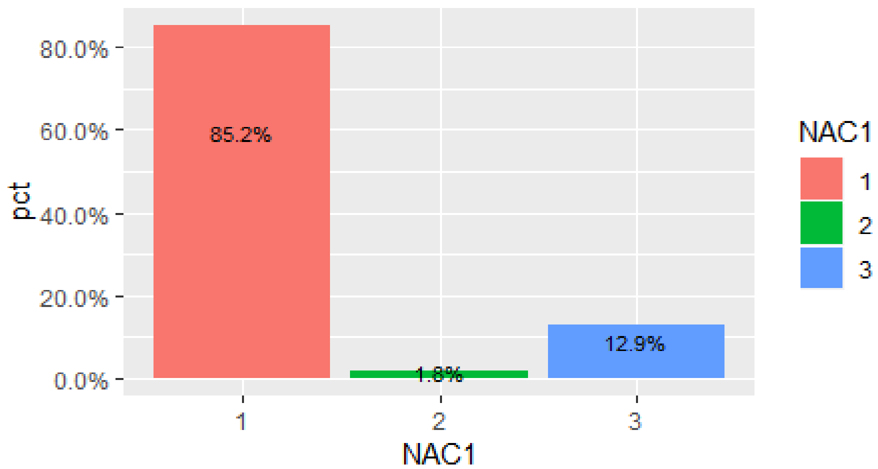
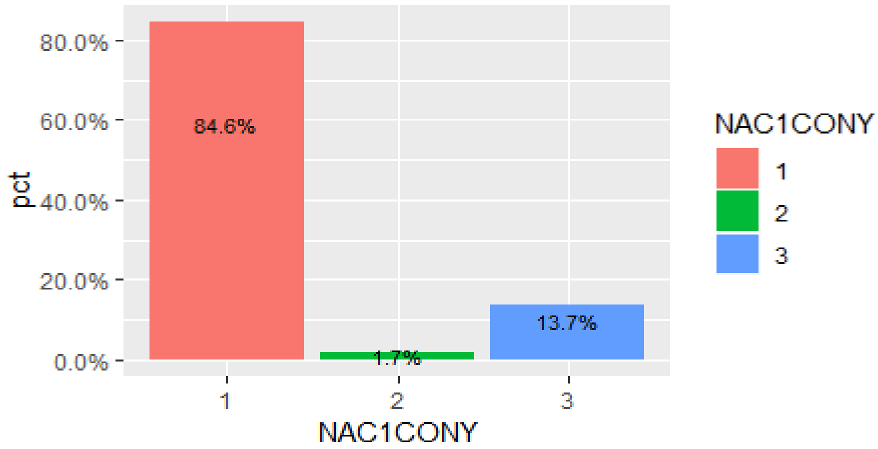


Figure 14 shows that 85.2% of the observations have the Spanish nationality and 12.9% of the observations have a foreigner status. Only 1.8% of the data have double nationality. The partner's version of this variable is also added as the immigrant situation of the partner might affect the decision of taking the parental permissions for both parents in the household. The variable is called NAC1CONY and its distribution is displayed in Figure 15.

Figure 15 Nationality Variable for the Partners



- *EDUCATION LEVEL (NFORMA in the microdata)*

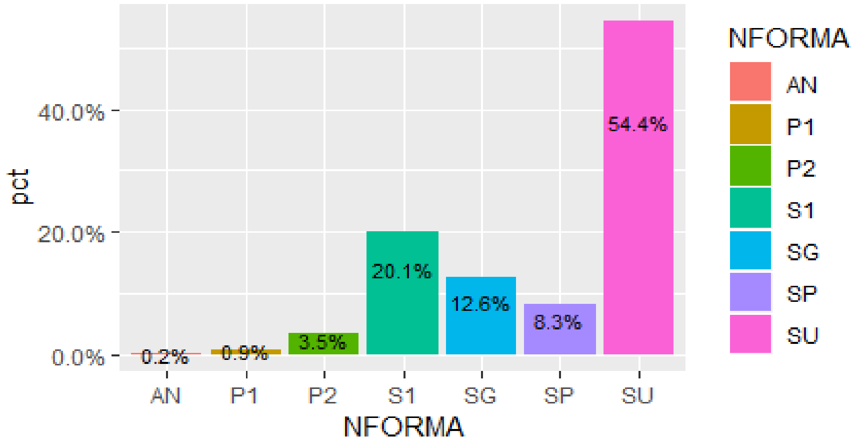
The education level can affect the analysis from different perspectives. On the one hand, males who have a high education level tend to support gender equality, therefore fathers who have higher education levels might be more conscious about balancing work and family life and sharing childcare responsibility. On the other hand, a person with a higher education level can have a better job, and therefore he/she could have more difficulties to use a permission because he/she faces more responsibilities and/or worries about losing job opportunities. Education level has 7 categories that are listed in the table below.

Table 5 Category Descriptions of Education Level

CODE	*DESCRIPTION*
AN	Illiterate
P1	Incomplete Primary Education
P2	Primary education
S1	First stage of secondary education
SG	Second stage of secondary education.
SP	Second stage of secondary education.
SU	Higher education

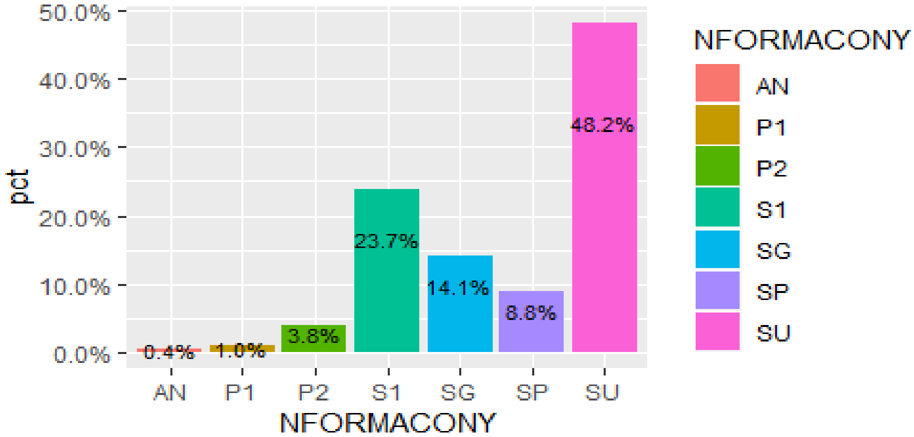
Figure 16 shows the distribution of the variable over the different categories. The category that has the highest percentage is higher education with 54.4% of the data. It is followed by the first stage of secondary education with 20.1% of the total observations in the sample.

Figure 16 Education Level



The partner's version of this variable is also added to the data sample in order to analyze whether the education level of partners affects the decision of taking a parental permission for both genders. The distribution of this variable is shown in Figure 17 below. The distribution among categories do not change much but the corresponding percentage of higher education is lower than in the original variable.

Figure 17 Education Level for the Partners



- *THE AGE WHEN FINISH THE HIGHEST LEVEL OF EDUCATION (EDADEST in the microdata)*

This variable represents the age of the individual when he/she reached the highest level of education on his/her academic life. It is an interval variable. It might have an impact given that a person who recently finishes his/her studies might be more unconfident about his/her job situation, thereby reducing the likelihood of taking a permission after a childbirth situation.

- *OCCUPATIONS (OCUP1 in the microdata)*

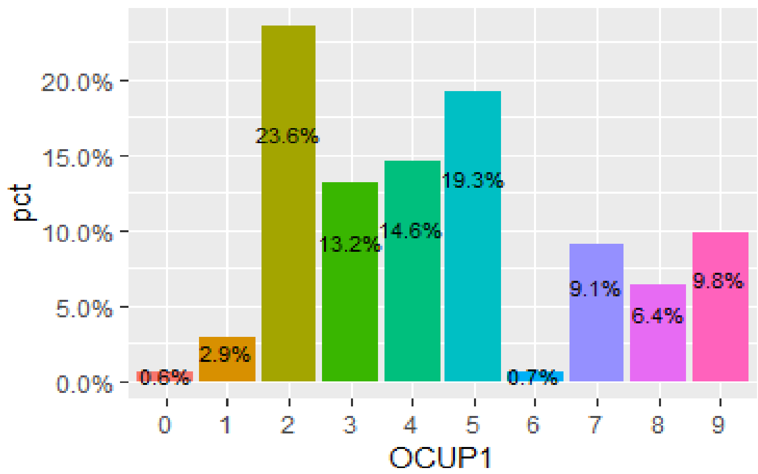
This variable represents the principal occupation of the interviewed people who is older than 16 years old and who has a work in the reference week. The variable has 10 categories which can be seen below with their descriptions.

Table 6 Category Descriptions of Occupation Variable

CODE	*DESCRIPTION*
0	Military occupations
1	Directors and Managers
2	Scientific and intellectual technicians and professionals
3	Support Technicians and Professionals
4	Accounting, administrative and other office employees
5	Catering, personal, protection and sales workers
6	Skilled workers in the agricultural, livestock, forestry and fisheries sector
7	Craftsmen and skilled workers in the manufacturing and construction industries
8	Plant and machinery operators and assemblers
9	No qualified occupations

In our sample, we distinguish four categories that have a higher percentage of frequency compared to the rest of categories. These categories are scientific and intellectual technicians and professionals, support technicians and professionals, accounting and administrative operators, and catering and sales workers.

Figure 18 Occupation Variable



The occupation variable's partner version has been created for adding the occupations of the individuals' partners in the sample. The original variable and its partners' version will be included into the statistical models in order to check wether belonging to a particular occupation group might have a significant impact on the decision of taking a parental permission.

- *PROFESSIONAL SITUATION (SITU in the microdata)*

This variable shows the professional situation of an interviewed person who is working in the reference week. It originally has 7 categories: business owner with employees; independent worker; a cooperative member; worker in a family business; wage earners in the public sector; wage earners in the private sector, and others. Table 7 shows a detailed description. During the data preparation process, it was decided that only wage earners from the public and the private sectors were going to be included in the data.

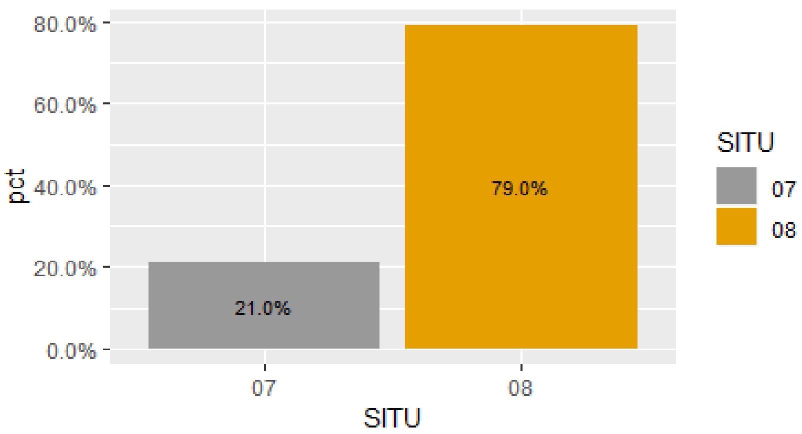
At the INE's webpage it is stated that "Wage earners with a determined duration job/contract are those employees whose main work will end after a predetermined period of time has elapsed, or after a period of time which is unknown from the outset" (INE, n.d.). Given that the working conditions of the other categories could be very different compared to those of a wage earner worker, data were filtered to just focus on wage earners. Therefore, only two categories of the variable were maintained in the sample.

Table 7 Category Descriptions of the Variable

CODE	*DESCRIPTION*
01	Businessman with employees
03	Freelance worker or employer without employees
05	Member of a cooperative
06	Help in the company or family business
07	**Wage earners in public sector**
08	**Wage earners in private sector**
09	Other situations

The distribution graph of the variable shows that most of the individuals in our data set are wage earners in private sector with the percentage of 79%. Only the 21% of the individuals are wage earners in the public sector.

Figure 19 Professional Situation Variable



Escot et al. (2013) showed that wage earners from the public sector are more likely to take a parental permission compared to wage earners from the private sector. We also include the partner's version of this variable to check whether the working conditions of the partners have an influence in the dependent variable.

- *ACTIVITY AREAS (ACT1 in the microdata)*

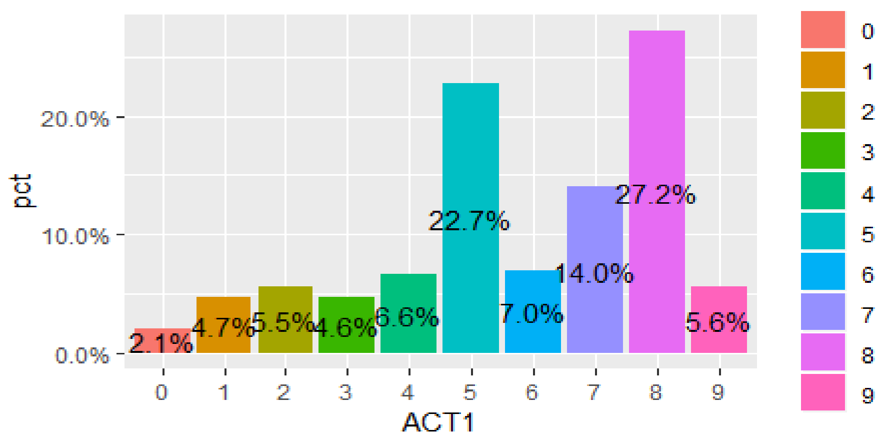
This variable specifies the activity area of an interviewed person who has a work in the reference week. Employment is distributed in 10 categories that cover the different employment sectors in the country. Table 8 contains a detailed description of each job category with its corresponding EAPS code.

Table 8 Category Descriptions of Activity Areas

CODE	*DESCRIPTION*
0	Agriculture, forestry and fishing
1	Food, textile, leather, wood and paper industry
2	Extractive industries, oil refining, chemical industry, pharmaceuticals, rubber and plastics industry, electric power supply, gas, steam and air conditioning, water supply, waste management.
3	Construction of machinery, electrical equipment and transport equipment. Industrial installation and repair
4	Constructions
5	Wholesale and retail trade, facilities and repairs. Car repair, hospitality.
6	Transportation and Storage Information and communications
7	Financial intermediation, insurance, real estate activities, professional, scientific, administrative and other services
8	Public Administration, education and health activities
9	Other services

As it can be seen in the graph, the categories which have the higher frequency rates are the area of public administration, education, health activities and the area of wholesale, retail trade, facilities and repairs.

Figure 20 Activity Areas

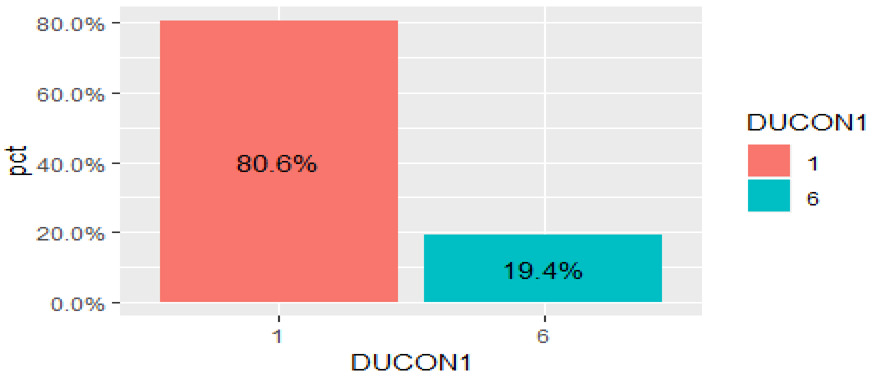


The variable's partner version has been added to our sample with the same aim that was explained in the previous selected variable.

- *TYPE OF THE CONTRACT (DUCON1 in the microdata)*

The type of contract is a binary variable that takes on a value of 1 if the contract is indefinite, and 6 if the contract is temporal. The survey question used for this variable was asked to all wage earner employees who had a job in the reference week. There might be a negative psychological effect associated with temporal contracts due to uncertainty. This might prevent fathers with this type of contract from using the paternity leave. Figure 21 shows that most of the individuals of our data set have indefinite contracts with a percentage of 80.6%.

Figure 21 Type of the Contract

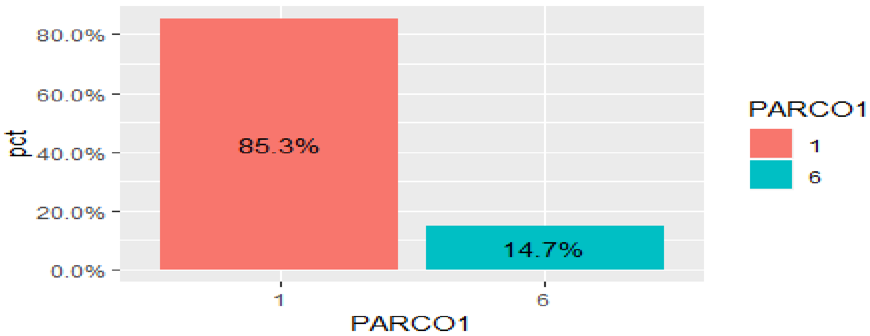


- *TYPE OF THE WORKING DAY (PARCO1 in the microdata)*

Another variable that is related with the job contract is the type of the working day. It gives information about whether an employee has a full-time or a part-time job. This variable is also important as one of the distinct characteristics of the job contract therefore it has been decided to include the variable and the version of the variable for the partners in the dataset for being able to analyze the working conditions of the parents in a more extensive way.

A full-time contract corresponds to a value of 1, while part-time contract corresponds to a value of 6. Figure 22 shows that most of the individuals have a complete working day contract with a percentage of 85.3%.

Figure 22 Type of Working Day



- *LABOUR STANDARTS (AOI in the microdata)*

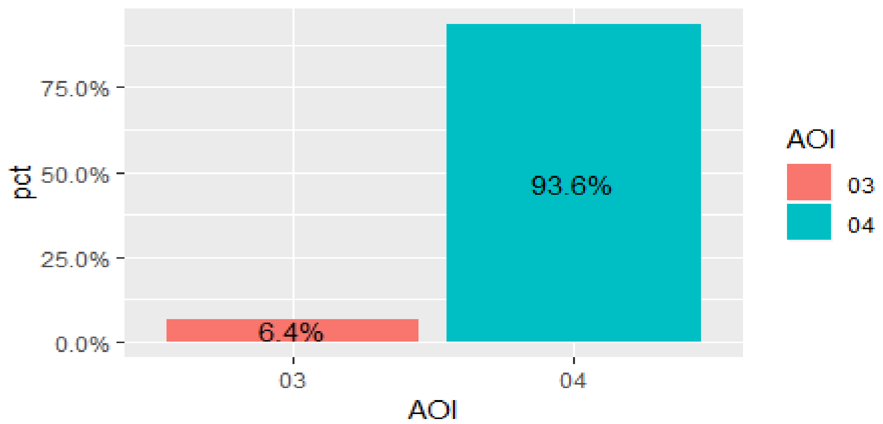
This variable represents a classification of respondents by relation to economic activity according to criteria from the International Labor Organization (ILO). The variable has 7 categories that are explained in Table 9.

Table 9 Category Descriptions of Labour Standards

CODE	*DESCRIPTION*
3	Employed underemployed due to insufficient hours
4	Rest of the employed
5	Unemployed looking for first job
6	Unemployed who have worked before
7	Inactive 1 (discouraged)
8	Inactive 2 (together with the discouraged form the potential assets)
9	Inactive 3 (rest of inactive)

When filtering our data to only contain wage earners who work either for the public or the private sector, we also filter the categories of this variable that correspond to individuals who either are unemployed or inactive according to ILO. As it can be seen in Figure 23, we ended up with two categories: "Employed underemployed due to insufficient hours" and "Rest of the employed".

Figure 23 Labour Standards

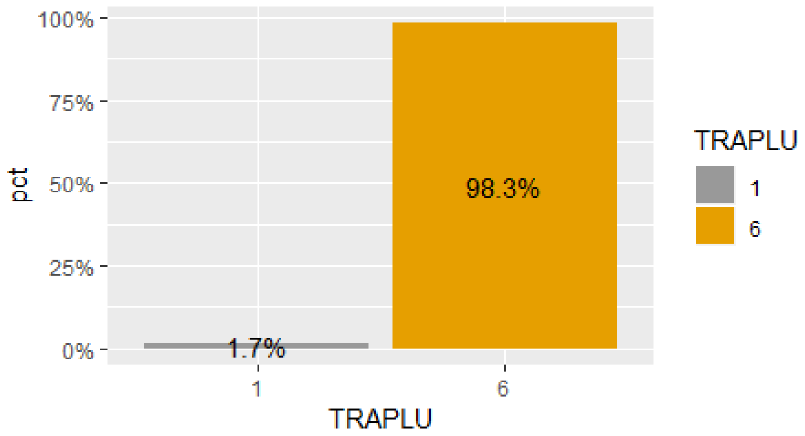


To have more detailed information on the working life of the partner, the partner's version of this variable was also included

- *SECOND JOB (TRAPLU in the microdata)*

The variable TRAPLU is a binary variable that represents whether an individual has more than one employment during the reference week when he/she is interviewed. If yes, the employment where the individual spends most of his/her working hours is considered as the principal employment and the other one is classified as secondary employment. Category 1 corresponds to an individual who has a second job, and the value 6 to an individual who has only one job. Usually, a second employment becomes necessary when there are economic problems in the household. A parental permission would therefore be less likely in this case. Figure 24 displays the observed frequencies.

Figure 24 Second Job



- *DURATION IN THE COMPANY (DUCOM in the microdata)*

It is an interval variable that shows the time period (in months) that an individual has been in his/her company. To predict the effect of this variable is difficult because, on the one hand, if an individual has not been working in the company for a long time, the use of a parental permission might be more difficult but, on the other hand, if an individual has been working for a long time in the company, he/she might hold a position of responsibility that might reduce the willingness to ask for a parental permission because the opportunity cost is higher.

- *TIME SINCE THE LATEST CONTRACT RENOVATION (DREN in the microdata)*

It is an interval variable that indicates the time duration by months since the time that an employee had the latest renovation of his/her contract. There is a high possibility that the variable might be correlated with the previous variable which indicates the duration in the company. We are going to investigate it in the following sections of SAS Miner.

- *ASSIGNMENT OF SOCIOECONOMIC STATUS (CSE in the microdata)*

This variable refers to the socioeconomic status of an interviewed person who was working during the reference week. The original 19 categories of the variable were reduced because the categories related to businessmen, freelancers and cooperative members were omitted, as we only include wage earners in our data sample. The categories considered are described in Table 10.

Table 10 Category Descriptions of Socioeconomic Status

CODE	*DESCRIPTION*
3	Directors and heads of farms
5	Other farm workers
10	Directors and managers of non-agricultural establishments, public administration management staff and members
11	Professionals, technicians and assimilated who carry out their activity on behalf of others
12	Professionals in exclusive occupations of public administration
13	Rest of administrative and commercial staff
14	Other services staff
15	Intermediate controls of non-agricultural establishments
16	Qualified and specialized operators of non-agricultural establishments
17	Operators without specialization of non-agricultural establishments
18	Armed Forces Professionals

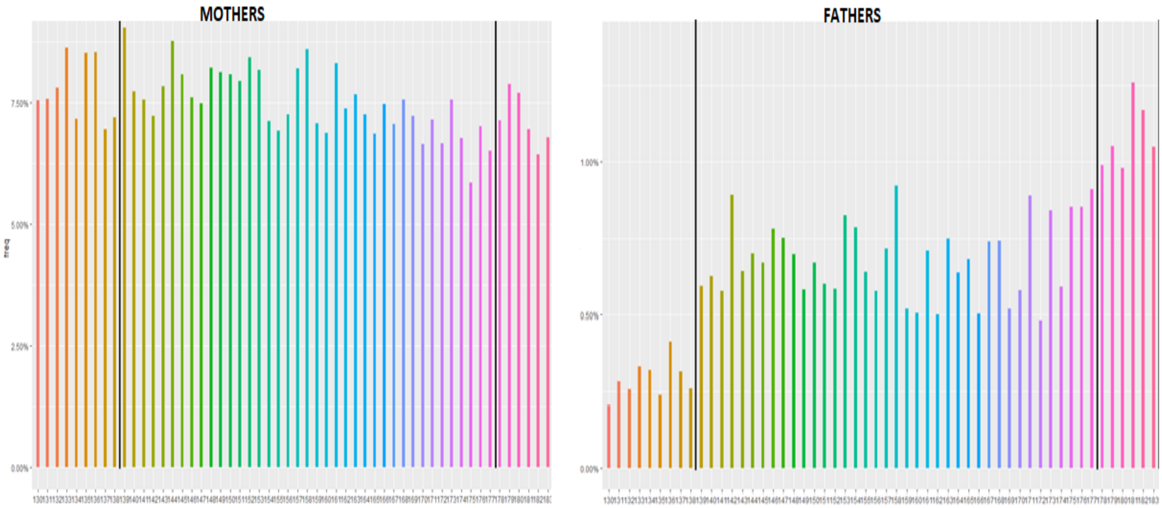
- *EMPLOYED PARTNER (TRAREMCONY in the microdata)*

The original variable (TRAREM) is corresponding the information if an individual is working in a paid job during the reference week. TRAREM, the original variable, is not included because our sample is composed by wage earners from the public and private sectors, and thus all individuals considered have an employment during the reference week of the survey. As the individuals' partners could be employment or unemployment, we are going to include the partners' variable in the data sample to see the impact of an employed partner in the decision of taking any parental permissions.

7.2 Comparison of take-up rates in both genders

Figure 25 offers a first look at the behavioral differences of both genders on taking a parental leave during the last 14 years. The graph on the left illustrates the females and the graph on the right the males. What we have done was creating a new graph where we can see the evaluation of the percentage of females on taking parental permissions as we created the same graph for the fathers in the data sample section. The graph has been divided with two vertical black lines in the same way with the graph of the fathers, in order to distinguish the times when the paternity reforms have been entered into force. In other words, these lines define three periods: no permission, 2 weeks of permission and 4 weeks of permission. After creating the new graph for the females, we have been put together the two graphs for mothers and fathers in order to observe the evolution of both genders in the take-up rate of parental permission.

Figure 25 Evaluation of Take-up Rates



The left graph shows that there is not a clear tendency among these three periods for mothers. As the paternity leave cannot be used by mothers, it is normal that we do not observe any effect of the reforms in the evolution of the take-up rate for the mothers. However, we observe a clear tendency for the fathers. From the first to the second period, we observe an increase in the use of the paternity permission. A positive slight change is also observed when comparing the second and the third periods.

This positive tendency for the fathers supports further statistical analysis. That is, we have to implement a battery of models to quantify the observed changes and to check whether they are representative.

8. Data Depuration in SAS Miner

After making the selection of variables from the survey, we transfer our data sample to SAS Enterprise Miner to carry out a data depuration process and to perform the statistical analysis. In Appendix C, it can be seen the selected variables with their roles in the data sample and measurement levels. The variable PERMISSION is assigned as the dependent variable and registered as binary.

The next step was running DMBD code for detecting if there are missing values and also for observing summary statistics of interval and categorical variables.

Figure 26 Interval Variables Summary Statistics

Variable	Label	Missing	N	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
DCOM		0	26682	0	480	91.8609	66.7399	0.78685	0.46204
DREN		0	26682	0	480	74.5360	64.8354	1.01650	0.92187
EDADEST		37	26645	7	54	20.8498	5.0192	0.75520	1.20350
NHIJOSLT5		0	26682	1	4	1.2655	0.4705	1.46234	1.20498

Figure 26 shows the four interval variables in the data set. EDADEST, which indicates the age when the latest education level was completed, is the only one with missing values. The limits of each variable were checked with the methods for detecting outliers, and atypical values were converted to missing values with the *Replacement Node*. We finally used the *Imputation Node* for imputing valid values to missing observations

Figure 27 Class Variables Summary Statistics

Variable	Label	Type	Number of Levels	Missing
ACT1		C	10	0
ACT1CONY		C	10	5105
AOI		C	2	0
AOICONY		C	7	0
CAMBIOPOLI		C	3	0
CCAA		C	19	0
CICLO		C	26	0
CSE		C	11	0
DUCON1		C	2	0
ECIV1		C	4	0
EDAD5		C	11	0
EDAD5CONY		C	11	0
FIRSTPERMISSION		C	2	0
NAC1		C	3	0
NAC1CONY		C	3	0
NFORMA		C	7	0
NFORMACONY		C	7	0
OCUP1		C	10	0
OCUP1CONY		C	10	5105
PARCO1		C	2	0
PARCO1CONY		C	2	5105
SECONDPERMISSION		C	2	0
SEXO1		C	2	0
SITU		C	2	0
SITUCONY		C	7	5105
TRAPLU		C	2	0
TRAREMCONY		C	2	0
permiso		C	2	0

Figure 27 shows the class variables. They are all free of missing values except from the following variables that we created for the partner: ACT1CONY, OCUP1CONY, PARCO1CONY and SITUCONY, which have the same amount of missing values. This is not a coincidence because these four variables were created in the same way from the INE survey. The procedure is as follows. The interviewed person is asked if he/she has worked or had a job during the reference week of the survey. If the answer is no, the questions that follow a yes answer are skipped. Therefore, these questions with no answer are recorded with a missing value. As the missing value actually means that a person does not have a job during the reference week, we created a new category of "UNEMPLOYED" for these 4 variables to keep these observations in the data set. As before, this process was implemented by the *Replacement Node*.

After the depuration process, we search for correlations between the interval variables. because multicollinearity between independent variables will weaken our statistical models. The Variable Correlation's Diagram was created using the *Variable Clustering Node*. Figure 28 shows the diagram.

Figure 28 Variable Correlations

We can see that there is a high correlation between DCOM (how long the individual has been in the company) and DREN (how long it has been since the latest contract in the company). In order to avoid possible estimation problems in the statistical models, we eliminated the variable DREN from the data set.

Sas Miner only allows independence between interval variables to be analyzed. To check for collinearity between the categorical variables we used the Variance Inflation Factor (VIF). The depurated data set was transferred to RStudio for running a specific coding program for the VIF calculation. If a variable has a high level of VIF (bigger than 4-5), it means that the variable has a high collinearity. We are concerned to analyze whether the following variables have collinearity between each other: principal occupation (OCUP1), principal activity (ACT1), professional situation (SITU), assignation of socioeconomic status (CSE) and the classification of occupied/active/unemployed (AOI). Figure 29 shows that CSE and OCUP1 have a high value of VIF and that is why the code removed the CSE variable from the list. As it is not good to have high collinearity between the independent variables, we eliminated CSE from the data set.

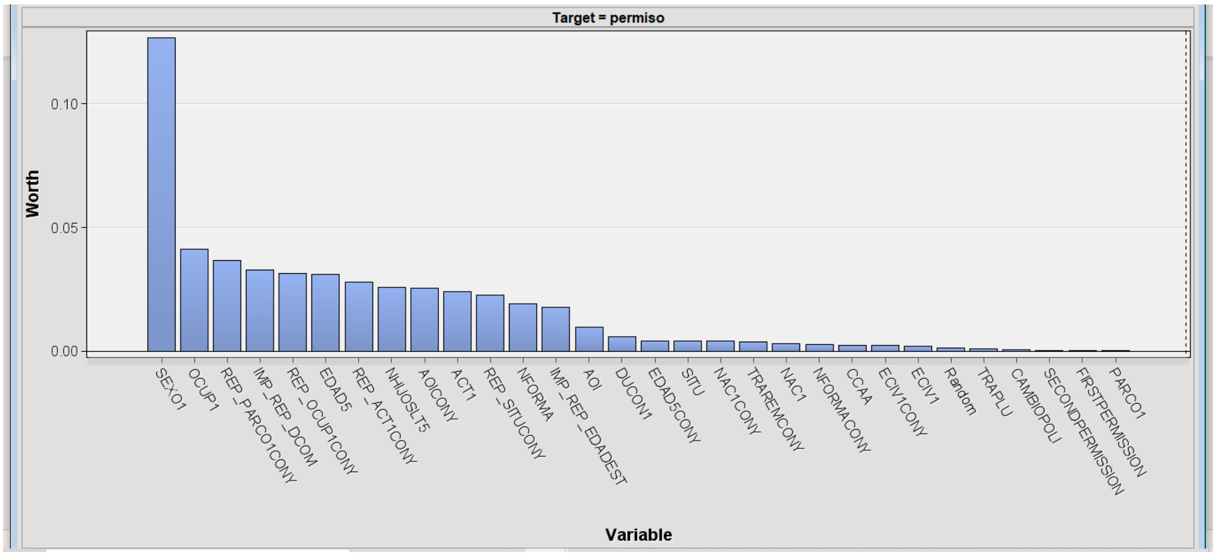
Figure 29 Variance Inflation Factor

```
  var                   vif
  AOI                   1.07642692059391
  ACT1                  1.97868798873099
  SITU                  2.6087880814753
  CSE                   7.61412466267898
  OCUP1                 7.47182934633631
  DUCON1                1.37791755556681
  permiso               1.21090681079721
  FIRSTPERMISSION       1.03844823077364
  SECONDPERMISSION      1.03599663168526
  EDAD5                 1.33679343548926
  IMP_REP_DCOM          4.30377899429268
  IMP_REP_DREN          4.34747852873979

removed:   CSE 7.614125
```

We next compare the worth of each variable using SAS Miner. First, we created a random variable with the node *Transformation of Variable*. Then, we put this variable and the other variables of the dataset into the graph Variable Worth executed by the node *StatExplore*. Figure 30 shows that the gender has the biggest influence on the permission variable. It is followed by occupation (OCUP1), the partner's working day (complete or partial) (PARCO1CONY) and the time duration of being in the company (DCOM). The random variable seems to have more importance than variables such as the second job (TRAPLU), the working day (PARCO1) and the variables that are related with the moment that the paternity reforms became applicable (CAMBIOPOLI, FIRST PERMISSION and SECOND PERMISSION).

Figure 30 Variable Worth



8.1 Regrouping Some Variables

Before statistical modelling, six variables were regrouped for a better interpretation of the output.

1.Age

Age (EDAD5) has 10 categories from 16 to 69 years old. We regrouped this variable into three categories. The first group consists of individuals from 16 to 24 years old, the second group is from 25 to 49 years old, and the last group is from 50 to 65 years old. We aimed at creating more interpretable groups that correspond to young, adult and elderly parents, respectively. We regrouped the partners' age variable (EDAD5CONY) in the same way.

2. Education

The variable NFORMA, which corresponds to the education level, has seven categories that range from illiterate to superior education. Superior education, which is the last level, is renamed as "High Education". Lower levels were joined together in a single category called "Lower Education". From this variable we could infer whether parents who have a high education level are more likely to use a parental leave.

3. Autonomous Communities

Escot et al. (2013) stressed that in some Autonomous Communities of Spain such as La Rioja, Navarra, Castilla-La Mancha, Murcia and Castilla y Leon, the regional government supports parents with some benefits while they are using a parental leave. According to that study, we regrouped the communities into two groups: encouraging communities and no encouraging communities. Thus, we reduced the nineteen Autonomous Communities of Spain in two categories according to the economic support policies that they are applying regarding parental leaves.

4. Principal Activity

The original variable of principal activity has ten categories that cover the different sectors of the Spanish labor market. In order to deal with rare categories and develop a better understanding of the variable, we used the *Interactive Grouping* node of SAS Miner for regrouping the ten categories. We ended up with the five groups listed in Table 11.

Table 11 New groups of Principal Activity

GROUP	*DESCRIPTION*
1	Agriculture, forestry and fishing / Constructions
2	Extractive industries / Construction of machinery, electrical equipment / Transportation and Storage Information and Communications
3	Food, textile, leather, wood and paper industry/ Other services
4	Wholesale and retail trade / Financial intermediation, insurance and real estate activities
5	Public Administration, education and health activities

5. Occupation

The occupation variable has initially ten categories. As before, we used the *Interactive Grouping* node to reduce the number of categories and to create groups with a similar event rate. The five groups created are presented in Table 12.

Table 12 New groups of Occupation

GROUP	*DESCRIPTION*
1	Scientific and intellectual technicians and professionals/ Accounting, administrative and other office employees
2	Support Technicians and Professionals/ Catering, personal, protection and trade sellers
3	Military occupations/ Directors and Managers
4	Plant and machinery operators and assemblers
5	Skilled workers in the agricultural, livestock, forestry and fisheries sector/ Craftsmen and skilled workers in the manufacturing and construction industries

6. Partner's Professional Situation

Professional situation has originally seven categories. As the data set was prepared with the parents that are wage earners, it remained with observations from only two categories of the variable: wage earners in the public sector and wage earners in the private sector. As it was stated in the variables' description section, the same variable was created for the partners of the individuals (SITUCONY). Due to the fact that the partners do not to have to be only in the categories of the wage earners, the partner´s professional situation variable has observations from different categories.

We applied Interactive grouping to SITUCONY, and the new four groups are described in Table 13.

Table 13 New groups of Partner´s Professional Situation

GROUP	*DESCRIPTION*
01	Other situations/Not Employed
02	Wage earners in public sector
03	Wage earners in private sector / Help in the company or family business
04	Businessman with employees /Freelance worker or employer without employees / Member of a cooperative

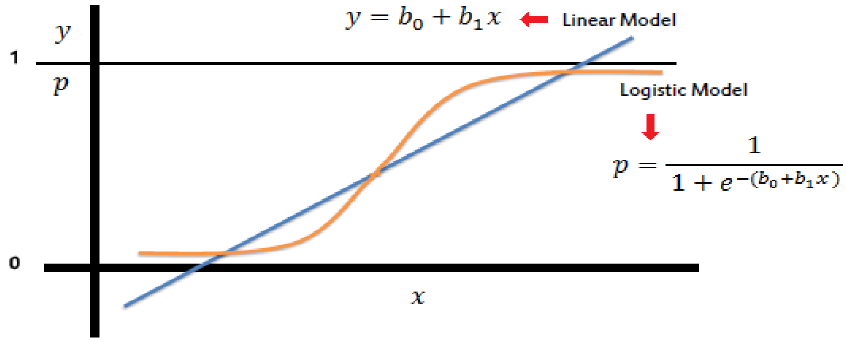
9. Methodology

As it was mentioned before, the purpose of the project is revealing in which conditions fathers use to take parental leaves. For this purpose, we are going to use the permission variable as our dependent variable for investigating which factors are affecting the parents in the decision of the permissions. As our target variable is a binary variable and we need to interpret and describe the relations between the independent variables and the dichotomous dependent variable, we are going to conduct to the Logistic Regression Analysis which is an appropriate statistical model for our project. Machine learning models are not used during the project as it is not possible to interpret the contributions of the variables in these models.

9.1 Logistic Regression

A logistic regression, also known as logit regression, is a statistical technique that predicts a dependent variable by examining the relationship between the target and the independent variables, which can be nominal, ordinal or continuous variables. The aim of the methodology is estimating the probabilities of events by considering relationships between features (Rouse, 2019). The main difference between linear regression and logistic regression is that linear regression analysis demands that the dependent variable is continuous. Besides, the linear regression requires a linear relationship between the dependent and independent variables when the logistic regression do not. In Figure 31, we can see that the logistic model creates a logistic curve while the linear model creates a lineal line.

Figure 31 Linear Model vs. Logistic Model. Source (Sayad, 2019)



Logistic regression is a derived version of a linear regression. The curve of the logistic model is built by taking into consideration the odds of the dependent variable not the probability as it is in the linear model. Below, the logistic regression equation is presented with odd ratios.

$$\log\left(\frac{p}{1-p}\right) = b0 + b_1x_1 + b_2x_2$$

Classification techniques are one of the main applications of data mining as many investigations try to reach out binomial outcomes such as Yes-No or 0-1 or True-False. Logistic Regression is a powerful method for predicting categorical dependent variables

9.1.1 Variable Selection Algorithms

While using logistic regression models, we applied three variable selection algorithms that are helpful for reducing the number of variables when some variables do not contribute to the dependent variable.

- ✓ Forward Selection: The method starts with no variable in the model. After that, the algorithm checks the p-value of all variables and chooses the variable that has the lowest p-value which is less than the reference p-value. The process continues one by one until the last significant variable is included in the model.
- ✓ Backward Selection: The algorithm starts with all variables and removes the variables one by one. At each iteration it is eliminated the variable that has the highest p-value which is greater than the reference p-value. The process continues until the last insignificant variable leaves the regression model.
- ✓ Stepwise Selection: The last algorithm is a combination of both forward and backward selection. The process starts as in the case of forward selection by adding the variables with lower p-values but later, those variables can be removed from the model if they lose their significance.

9.1.2 Information Criteria

We applied different information criteria for each variable selection algorithm in SAS Miner. Information criteria help to compare models. The lower the value, the better the model. The three criteria consider the model error and a penalty score according to the number of parameters of the model. They differ on how this penalty score is calculated. The formulas of the criteria can be seen below.

- AIC (Akaike information criterion): n ln(SSE n) + 2p
- BIC (Bayesian information criterion): n ln(SSE n) + 2q(p + 2 − q)
- SBC (Schwarz information criterion): n ln(SSE n) + p ln(n)

9.1.3 Comparison Criteria for Regression Models

In the step of comparing our models, we consulted two types of comparison criteria.

- ❖ Misclassification rate: This rate is calculated by summing up the false classified incidents (false positives and false negatives) and dividing them by the total number of incidents. The lower the rate, the better the model.
- ❖ Receiver Operating Characteristic Curve (ROC): The ROC curve is a graphical representation of the sensitivity versus specificity for a binary classifier system. In other words, it is a comparison of True Positives and False Positives. The area which stays below the curve is used as a comparison criterion of the logistic regression models. When the area is bigger, the model is better.

9.1.4 Difference in Differences Estimation

Difference in differences (DID) is a statistical method used in econometrics to study causal effects. The differential effect of a treatment is estimated by comparing the diversity of outcomes between a treatment group that is affected by the treatment, and a control group that is not affected. The treatment is the paternity leave reform. Thus, we have two treatments. The treatment group is composed by the fathers, who are influenced by these reforms, and the control groups is formed by the mothers, who are not influenced. See Angrist and Pischke (2009) for further details on this technique.

9.1.5 Data Partition

Before starting to create the regression models, we have split the data set into two subsets.

- ✓ Training set: This subset is created for training a model. 70% of the dataset was converted to the training set.
- ✓ Test set: This subset is created for testing a trained model. 30% of the dataset was converted to the test set.

9.1.6 Confusion Matrix

Confusion matrix is a table that outlines the performance of a classification model. The matrix describes the performance on the test set that has 30% of the data. It gives a summary of prediction results by dividing them into four parts.

- True Positive (TP): When the observation is positive, and the prediction is positive by the model. In other words, correct positive prediction.
- False Negative (FN): When the observation is positive, but the prediction is negative by the model. In other words, incorrect negative prediction.
- True Negative (TN): When the observation is negative, and the prediction is negative by the model. In other words, correct negative prediction.

- False Positive (FP): When the observation is negative, but the prediction is positive by the model. In other words, incorrect positive prediction.

From the confusion matrix, some measures for the performance of the model can be generated.

Accuracy: This measure is calculated as the total number of all the correct predictions divided by the number of the dataset. In case of a perfect model, the accuracy rate is 1.

Sensitivity (Recall): This measure is calculated as the number of correct positive predictions divided by the total number of positive observations. If a model is perfect, the sensitivity rate is 1.

Specificity (True Negative Rate): This measure is calculated as the number of correct negative predictions divided by the total number of negative observations. If a model is perfect, the specificity rate is 1.

Precision (Positive Predictive Value): This measure is calculated as the number of correct positive predictions divided by the total number of positive predictions. In case of a perfect model, the precision rate is 1.

10. Models

10.1 Benchmark model with gender and the derived variables from paternity permission

Before including all the selected variables in the logistic regression, we estimated an initial model with just three variables that served as benchmark. These variables are a gender variable and two variables that were derived from the permission variable to indicate the period that follows the introduction of the reform: FIRSTPERMISSION for the first reform and SECONDPERMISSION for the second reform. With this model, we wanted to check whether fathers are affected by these reforms when taking the paternity permission. In other words, we wanted to analyze if paternity reforms have a casual effect. If there are two different groups in a sample that are assigned to a treatment and a control group, and if these groups can be differentiated after a treatment, it can be said that the treatment is causing a difference. In our data sample, the fathers are the treatment group and the mothers are the control groups, as the paternity leave reforms are implemented for male wage earners after a childbirth. In order to analyze the impact of the reforms on fathers, we are going to create interactions between the variables of the reforms and the gender variable (with the male character of the variable).

We estimated this model with the data sample (26.886 observations) that we created with the under-sampling method. But before estimation, some actions were taken:

- ❖ Data partition was performed. The database was divided into training and test. The training data have 70% of total data and the test data have the remaining 30%.

- Two interaction variables were created: the first interaction is between the first reform and the male character, and the second interaction is between the second reform and males.

The results of the benchmark model are displayed in Table 14.

Table 14 Benchmark Model

	ESTIMATE	STD.ERROR	Z VALUE	Pr(>IzI)	Exp
(Intercept)	0.83778	0.05137	16.310	<2e-16	2.311
GENDER:					
MALE	-3.40530	0.13036	-26.126	<2e-16	0.033
FIRSTPERMISSION:					
YES	-0.07937	0.05605	-1.416	0.157	0.923
SECOND PERMISSION:					
YES	0.05920	0.6547	0.904	0.366	1.06
MALE:FIRSTPERMISSION.YES	0.95039	0.13827	6.874	6.26e-12	2.586
MALE:SECONDPERMISSION.YES	0.55411	0.11551	4.797	1.61e-10	1.740

According to the results in Table 14, the variable GENDER has a p-value smaller than 0.001, therefore it is significant at 0.1% level. Regarding to the estimated coefficient of the variable, we can say that fathers are less likely to use a parental permission. The first permission and the second permission variables are not significant for the models as their p-values are bigger than 0.001. But these variables' interactions with gender are significant as their p-values are less than 0.001. We expected these results because the reforms only affect the fathers. Therefore, it is normal that the reform variables become significant when they interact with males. According to the estimated values of the coefficients for the two interactions, we can say that fathers are more likely to use a parental permission when a reform is implemented.

We now focus on the exponential values of the estimated coefficients to interpret them. Gender has an odds ratio of 0.03 with its category male. It means that men are 97% less likely to use a permission than women. The interaction between males and the first permission has an odds ratio of 2.586. This interaction compares the effect of first permission for males who were examined after the first reform came into force and those males who were not. As the odds ratio is 2.586, it indicates that males who were surveyed after the first permission reform are 2.58 times more likely to use a permission that other males who were not in that period. The second interaction between males and the second permission reform has an odds ratio of 1.740. Therefore, we can say that the males who were in the second permission time period are 1.74 times more likely to use a permission compared to males who were not in that period.

10.2 Logistic Regression Models

We next put all the selected variables into *Regression* nodes to create different models. Three main paths for connecting Regression nodes were considered. They are listed below.

- The first path was to create the regression models just after the data partition without doing other process in between.
- The second path was to use the *Variable Selection* node between the data partition and the regression models in order to execute the models with an additional step of selection of variables by the node.
- The third path is to use the *Decision Tree* node between the data partition and the regression models in order to try another way of selection of variables.

The *Variable Clustering* node was discarded because we have few interval variables in the dataset. The *Transform Variables* node was not used neither for not losing the interpretability of the variables.

For each path, we run a default model and a battery of successive regression models using the Forward, Backward and Stepwise selection methodologies. We tried the different selection criteria that are integrated in the *Regression* nodes of SAS Miner for variable selection. The criteria are AIC, SBC and Cross Validation Misclassification.

All the models created are compared according to the Misclassification Rate in Test. We used the *Comparison Model* node for that. The models that are created without applying either the *Decision Tree* or the *Variable Selection* node have lower misclassification rates. Figure 32 shows that the first 10 models with lower rates are the ones that are created from the first path. The Default Model where neither a variable selection algorithm nor selection criteria were applied ranked in tenth place. The first 10 are then followed by two models from the third path: the first one is with the Forward algorithm and the second one is with the Stepwise Algorithm, and both models with SBC criteria. A model from the second path (selection of variables applied before regression models with a Stepwise algorithm and a Cross Validation Misclassification criterion) closes the table.

Figure 32 Misclassification Rate of the Models

Model Description	Probar: tasa de clasificación errónea ▲
Backward.AIC	0.225081
Backward.CVM	0.225081
Backward.SBC	0.225331
Forward.AIC	0.225831
Stepwise.AIC	0.225831
Stepwise.CVM	0.22708
Forward.CVM	0.22708
Stepwise.SBC	0.227954
Forward.SBC	0.227954
DefaultModel	0.228079
Tree/Forward.SBC	0.230577
Tree/Stepwise.SBC	0.230577
Select/Stepwise.CVM	0.230952

Figure 33 shows a table for comparing the selected variables from each model using the five best models from the previous step.

Figure 33 Best Models and Their Variables

Backward.AIC	Backward.SBC	Backward.CVM	Stepwise.AIC	Forward.AIC
ACT1	ACT1	ACT1	ACT1	ACT1
CAMBIOPOLI	CAMBIOPOLI	CAMBIOPOLI	DUCON1	DUCON1
DUCON1	DUCON1	DUCON1	DCOM	DCOM
ECIV1	ECIV1	ECIV1	EDADEST	EDADEST
DCOM	DCOM	DCOM	NAC1	NAC1
EDADEST	EDADEST	NFORMA	NHIJOSLT5	NHIJOSLT5
NAC1	NAC1	NFORMACONY	OCUP1	OCUP1
NHIJOSLT5	NHIJOSLT5	PARCO1CONY	PARCO1	PARCO1
OCUP1	OCUP1	SITUCONY	NFORMA	NFORMA
PARCO1	PARCO1	SEXO1	NFORMACONY	NFORMACONY
EDAD5CONY	NFORMA	SITU	SITUCONY	SITUCONY
NFORMA	NFORMACONY	TRAREMCONY	SECONDPERMISSION	SECONDPERMISSION
NFORMACONY	SITUCONY		SEXO1	SEXO1
PARCO1CONY	SEXO1		SITU	SITU
SITUCONY	SITU		TRAREMCONY	TRAREMCONY
SEXO1	TRAREMCONY			
SITU				
TRAREMCONY				

There are 9 variables that are selected from all five models. These variables are *ACT1* (principal activity), *DCOM* (time period of being in the company), *DUCON1* (type of the contract), *NFORMA* (education level), *NFORMACONY* (education level of the partner), *TRAREMCONY* (employed partner), *SEXO1* (gender), *SITU* (professional situation), *SITUCONY* (professional situation of the partner). There are 5 variables that are selected from four out of five models. These variables are *EDADEST* (age at the highest level of education), *NAC1* (nationality), *NHIJOSLT5* (number of children less than 5 years old), *OCUP1* (occupation) and *PARCO1* (type of working hours).

For the rest of this section, we organize the work as follows. First, we choose the best regression model among the five selected models. Secondly, we discuss the need to add other variables that are not in the best regression model, but they could be helpful for a better evaluation of the take up rates of the permissions. Lastly, we interact the gender variable with each of the selected variables to check whether they continue to be significant when interacted. These interactions aim at identifying the variables that could affect the father's decision of using the parental permission, and to check whether there are differences in the behavior of the target group (fathers) and the control group (mothers). In addition, it was created decision trees with all the variables before applying the final regression model in order to see the trees' branches which give an idea of possible relations among variables for the classification. After observing the results, it was seen that all the decision trees were starting to be split with the gender variable for classifying whether a permission was taken or not. Consequently, it was concluded that interactions with gender variable might be beneficial to increase the prediction ability of the final regression model.

We took the five best models and put them in a repeating process of Training-Test as it is showed in Appendix D. In the *Start Groups* node, the *Index* node was selected for determining how many times the loop would be processed through the process flow diagram. The number of repetitions was chosen to be 15 times. We repeated the process of Training-Test with random selected seeds to ensure that the models have enough variability in their performances.

After finishing the Training-Test process, we created a boxplot based on the Misclassification rate of the five final models. Results are displayed in Figure 34. The results are similar but the model of the Backward algorithm with the AIC criterion is the one that has highest variance. Therefore, it is less trustable compared to the others. We also observe that the models Forward algorithm with AIC criterion and Stepwise algorithm with AIC criterion have exactly the same results. Moreover, these models have lower variance compared to the others, and the average of the misclassification rate is almost the same for all models. Consequently, the Stepwise-AIC and the Forward-AIC models outperformed the rest of the models. Figure 33 shows that these models selected the same variables. Besides, their Test ROX Index had the same value. Therefore, we concluded that these models are essentially the same, and only the results of the Forward-AIC model are further analyzed.

Figure 34 Boxplot Result of Misclassification Rates

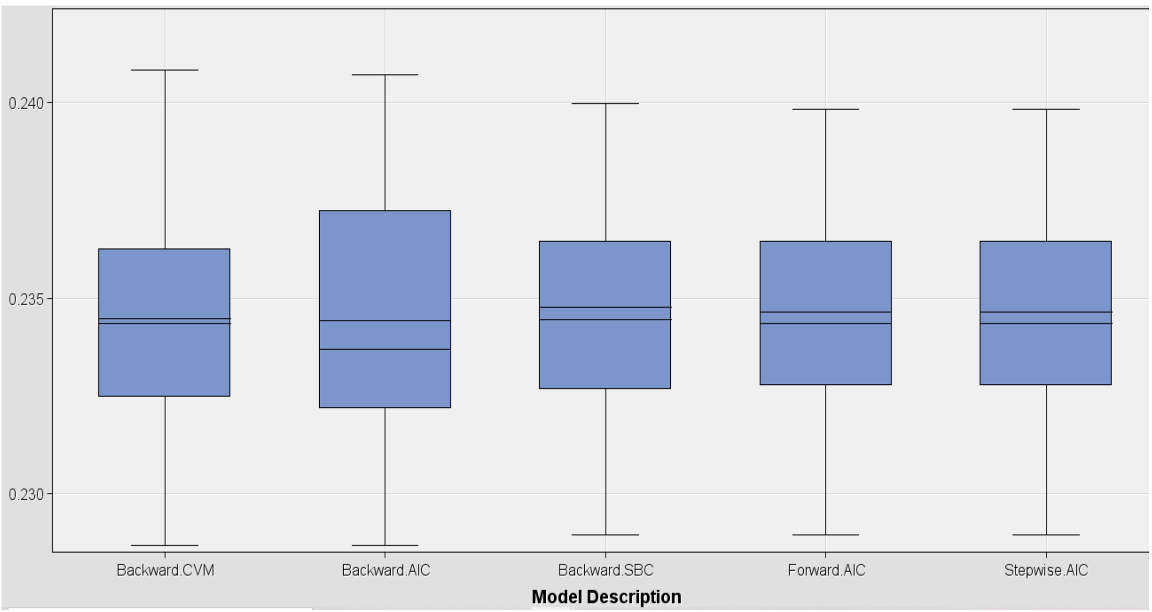


Table 15 lists the Analysis of Effects for each variable included in the Forward-AIC model. The Type 3 Analysis of Effect is based on the value of the Wald Chi-Square test statistic, which is an indicator that shows whether a variable is significant for the model or not. All the listed variables are significant for the model as their p-values are less than 0.05 (5% significance level). According to the Wald scores of the variables, gender (SEXO1) has the highest value (2535.10). This means that it is the variable that contributes the most to the regression model. After that, the variables NHIJOSLT5 (number of children less than 5 years old) and TRAREMCONY (employed partner) have also an important contribution to the model with high values of the Wald score. However, PARCO1 (type of working hours), DUCON1 (type of the contract), DCOM (time period of being in the company) and ECIV1 (civil status) have much lower Wald scores compared to the first three variable, despite this they also have considerable importance for the model.

Table 15 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
ACT1	9	23.0474	0.0061
DUCON1	1	164.4774	<.0001
IMP_REP_DCOM	1	112.2788	<.0001
IMP_REP_EDADEST	1	4.0563	0.0440
NAC1	2	9.9775	0.0068
NHIJOSLT5	1	688.9287	<.0001
OCUP1	9	59.2723	<.0001
PARCO1	1	262.2379	<.0001
REP_NFORMA	1	7.7147	0.0055
REP_NFORMACONY	1	5.7789	0.0162
REP_SITUCONY	6	19.0138	0.0041
SECONDPERMISSION	1	14.1495	0.0002
SEXO1	1	2535.1028	<.0001
SITU	1	7.3033	0.0069
TRAREMCONY	1	667.5420	<.0001

In the next part, we took the best model, the Forward-AIC, as our reference model, and we added some new variables, which were considered as important for our research question. The best model has the nine variables that were selected by the five final models and it also has the five variables that were selected from four out of the 5 final models. That is why we did not have to include these variables manually. We added as extra variables, EDAD5 (age), CCAA (autonomous communities) and FIRSTPERMISSION (first reform) to analyze whether they have an impact on the decision of taking the parental permission. Finally, we created interactions between the gender and the rest of the variables to assess potential differences between the target group (fathers) and the control group (mothers). Besides, the estimation of these interactions will help us to find out the determinants of the fathers in the use of parental permission. Table 16 shows the results of the logistic regression model. The variables were written with their descriptions instead of their original names in order to facilitate interpretation. The SAS Miner results with the original names of the variables can be seen in Appendix E.

Table 16 Final Model

	ESTIMATE	STD.ERROR	Z VALUE	Pr(>IzI)	Exp
(Intercept)	-1.6106	0.2209	52.70	<.0001	0.200
Contract Type:					
Indefinite Contract	0.2785	0.0346	64.76	<.0001	1.321
FirstPermission:					
BeforeFirstPermission	-0.179	0.0383	21.83	<.0001	0.836
Principal Activity:					
Group1	-0.0795	0.0732	1.18	0.2776	0.924
Group2	0.0802	0.0517	2.41	0.1205	1.084
Group3	-0.0218	0.0699	0.1	0.7548	0.978
Group4	0.0791	0.0492	2.58	0.1081	1.082
Occupation:					
Group1	0.0163	0.0553	0.09	0.7686	1.016
Group2	0.0837	0.0476	3.09	0.0788	1.087
Group3	-0.3128	0.0612	26.1	<.0001	0.731
Group4	0.0162	0.0782	0.04	0.8362	1.016
PartnerProf.Situ.:					
Group1	-1.0845	0.0555	381.4	<.0001	0.338
Group2	0.1007	0.053	3.61	0.0576	1.106
Group3	0.2362	0.04	34.95	<.0001	1.266
TimeperiodinJob(month)	-0.00324	0.00038	72.73	<.0001	0.997
Ageinthehighesteducation	-0.00812	0.00569	2.03	0.1537	0.992
Nationality:					
Spanish	0.0316	0.0713	0.2	0.6582	1.032
DoubleNationality	0.2084	0.125	2.78	0.0956	1.232
Numberofchildren	1.0538	0.0488	467.11	<.0001	2.869
WorkinghoursinJob:					
Fulltime	0.3539	0.0726	23.79	<.0001	1.425
Regions:					
Encouraging region	0.0505	0.025	4.09	0.0432	1.052
Age:					
Age(16-24)	0.1456	0.2043	0.51	0.476	1.157
Age(25-49)	-0.1022	0.1442	0.5	0.4783	0.903
LevelofEducation:					
HighEducation	0.0668	0.0317	4.44	0.0352	1.069
PartnerLevelofEducation:					
HighEducation	0.0492	0.0272	3.27	0.0704	1.05
SecondPermission:					
BeforeSecondPermission	-0.1953	0.0347	31.63	<.0001	0.823
Gender:					
Male	-0.9981	0.1936	26.57	<.0001	0.369
ProfessionalSituation:					
WageEarnerPublicSector	0.2458	0.0512	23.01	<.0001	1.279

EmployedPartner:					
Partnerwithjob	-0.9808	0.0359	748.49	<.0001	0.375
Contract Type*Gender:					
*Indefinite Contract*Male*	-0.0848	0.0346	6.01	0.0142	0.919
FirstPermission*Gender:					
*BeforeFirstPermission*Male*	-0.2267	0.0383	35.03	<.0001	0.797
Principal Activity*Gender:					
*Group1*Male*	0.0208	0.0732	0.08	0.7767	1.021
*Group2*Male*	0.1123	0.0517	4.72	0.0298	1.119
*Group3*Male*	0.0364	0.0699	0.27	0.6025	1.037
*Group4*Male*	-0.019	0.0492	0.15	0.6995	0.981
Occupation*Gender:					
*Group1*Male*	-0.00067	0.0551	0	0.9903	0.999
*Group2*Male*	0.0198	0.0476	0.17	0.6768	1.02
*Group3*Male*	0.0247	0.0612	0.16	0.687	1.025
*Group4*Male*	-0.1034	0.0781	1.75	0.1857	0.902
PartnerProf.Situ.*Gender:					
*Group1*Male*	-0.6214	0.0555	125.37	<.0001	0.537
*Group2*Male*	-0.00044	0.053	0	0.9933	1
*Group3*Male*	0.0369	0.04	0.85	0.3556	1.038
TimeperiodinJob(month)*Male	0.000274	0.00038	0.52	0.4698	1
Ageinthehighesteducation*Male	0.00623	0.00569	1.20	0.2740	1.006
Nationality*Gender:					
*Spanish*Male*	-0.08	0.0713	1.26	0.2617	0.923
*DoubleNationality*Male*	0.2064	0.125	2.73	0.0988	1.229
Numberofchildren*Male	-0.1304	0.0487	7.16	0.0075	0.878
WorkinghoursinJob*Gender:					
*Fulltime*Male*	-0.0465	0.0726	0.41	0.5217	0.955
Regions*Gender:					
*Encouragingregion*Male*	0.021	0.025	0.71	0.4	1.021
Age*Gender:					
*Age(16-24)*Male*	-0.2546	0.2042	1.55	0.2126	0.775
*Age(25-49)*Male*	-0.0996	0.1442	0.48	0.4897	0.905
LevelofEducation*Gender:					
*HighEducation*Male*	-0.0219	0.0287	0.59	0.4441	0.978
PartnerLevelofEducation*Gender					
*HighEducation*Male*	0.00198	0.0271	0.01	0.9417	1.002
SecondPermission*Gender:					
*BeforeSecondPermission*Male*	-0.1578	0.0347	20.7	<.0001	0.854
Prof.Situ.*Gender:					
*WageEarnerPublicSector*Male*	0.208	0.0512	16.48	<.0001	1.231
EmployedPartner*Gender:					
*Partnerwithjob*Male*	-0.6931	0.0358	373.98	<.0001	0.5

The variable Contract Type is statistically significant (p-value < 0.0001) and its odds ratio is 1.321. It means that parents who have an indefinite contract are 1.32 times more likely to use a childbirth permission. The interaction between indefinite contract and male is significant because its p-value is less than 0.05. Its Wald Chi-Square value is 6.01, therefore it contributes to the model. However, without the interaction, the variable contract type has a positive effect on taking paternity leave, but the effect is negative with the interaction. According to the interaction's odds ratio (0.91), men who have an indefinite contract are 0.09 times less likely to use a parental permission compared to men who have a temporal contract.

The variable First Permission has a significant effect on the dependent variable. As its odds ratio is 0.836, before the introduction of the first policy, both men and women are 16% less likely to use a parental permission compared to the period after the first policy. The interaction between first permission and male is significant for the model as well, and men before the first paternity reform are 20% less likely to use the parental permission. Since the paternity reform was only implemented to the fathers, it is normal that the interaction has a lower odds ratio than the odds ratio of the variable without interaction. Regarding this result, the fathers´ take-up rate has changed more compared to the take-up rates of men and women when the policy come into force.

For the variable Principal Activity, group 5 is excluded as the reference group for making comparisons with the rest of the groups. The four groups of the variable are not significant at the 5% as their p-values are higher than 0.05. The interactions with male follow the same pattern except for the interaction of Group 2 and Male, which has a p-value less than 0.05. The interaction's odds ratio is 1.119, therefore men from Group 2 (Food, Textile, Leather, Extractive Industries and Transportation sectors) are 11% more likely to use parental permission compared to men from Group 5 (Public Administration and Education).

For the variable Occupation, the group 5 is excluded from the model. Occupations from Groups 1, 2 and 4 are not significant at the 5%, as their p-values are higher than 0.05. Group 3 is significant (p-value<.0001) and its odds ratio is 0.731, which indicates that parents from Group 3 (Military services, Directors) are 26% less likely to use the parental permission compared to the people from Group 5 (Installation Operators). Regarding the interactions of the four groups with the Male character, none of them are significant.

For the variable Professional Situation of the Partner, Group 4 is excluded from the model. Groups 1 and 3 are significant for the model. Men and women whose partner is from Group 1 (Unemployed and other situations) are 66% less likely to use the parental permissions compared to people whose partner is from Group 4 (Businessmen, independent worker and cooperative member). It seems logical, as parents with one of its members at home might make less use of the parental leave. Men and women whose partner belong to Group 3 (Private sector and family business) are 26% more likely to use the parental permissions compared to people whose partners are from Group 4. About the interactions with Male character, only the interaction of Group 1 and Male is significant. Men with partners from Group 1 (Unemployed and other situations) are 46% less likely to use the parental permission

The continuous variable Time Period in Job, which indicates how many months the individual has been in the company, is significant at the 5% with a p-value less than 0.0001. Its odds ratio is 0.997. This means that when the value of time period increases by 1, both men and women are 0.3% less likely to take the parental leave. The interaction between this variable and the male category is not significant.

The other continuous variable Age in the Highest Education is not significant as the p-value is bigger than 0.05. The interaction of the variable with the fathers do not have a statistical significance neither for the model.

For the Nationality variable, the category Foreigners is excluded by the model. The categories of Spanish nationality and Double nationality are not significant. The interactions with the Male character are not significant either, as their p-values are higher than 0.05.

Another continuous variable, the Number of Children less than 5 Years Old, is significant at the 5% level. According to the odds ratio (2.869), when the number of children increases by 1, parents (men/women) are 186% more likely to use the parental permission. The interaction of the variable with Male is still significant although with a lower significance rate and with an opposite estimation effect. The odds ratio is 0.878, therefore when the number of children increases by 1, parents are 22% less likely to use the parental permissions.

The categorical variable Working Hours in Job, which indicates whether the contract is full-time or part-time, is considered as significant variable for the model .The parents (men/women) who have a full-time contract are 42% more likely to use the parental permission than parents who have a part-time contract. The interaction between full-time contract and male is not statistically significant for the model.

The Autonomous Communities were grouped into two groups according to their regional policies. The category of encouraging regions is significant at the 5% level. Parents who live in encouraging regions are 5% more likely to use the permission. The interaction of the encouraging regions and Male is significant as well, thus fathers who live in encouraging regions are 2% more likely to use the permission.

For the Age variable, the age group between 50 and 65 years old was excluded from the model. The groups of 16-24 years old and 25-49 years old are not significant, which means that there is not a statistically difference between the age groups. The interactions of the fathers with these age groups are not significant either.

Having a high education level has a positive and significant effect for the parents, both men and women. A person who has a high education level is 6% more likely to use the parental permission. The interaction for the fathers is not statistically significant. Having a partner who has a high education level is not statistically significant as the p-value is higher than 0.05. The interaction of the variable with the fathers is not significant either.

The variable Second Permission, which indicates the second policy of paternity leave, has a significant effect on the dependent variable. As its odds ratio is 0.823, before the second policy, men and women are 17% less likely to use the parental permission compared to the period that is after the second policy. The interaction between the second permission and the fathers is significant for the model as well, and thus men before the second paternity reform are 15% less likely to use the parental permission.

The Gender is significant at the 5% level. According to its odds ratio (0.369), men are 61% less likely to take the parental permission than women. As we analyzed the gap between mothers and fathers on taking a parental leave in previous sections, we can say that the result has a logical point.

The categorical variable of Professional Situation, which shows whether the individual is a wage earner in the public sector or the private sector, is statistically significant at the 5%. Women and men who are wage earners in the public sector are 27% more likely to use a parental leave than people in the private sector. The interaction of Professional Situation with the father is significant as well and has a similar effect. The fathers who are in the public sector are 23% more likely to use a parental leave.

The binary variable Employed Partner is statistically significant with a high value of Wild Chi-square. According to its odds ratio, men and women who have employed partners are 62% less likely to take a parental leave. The interaction with the Male character has its contribution to the model as it is significant. The fathers who have employed partners are 50% less likely to take a parental leave.

After analyzing the parameter estimation of the final model, an analysis of its performance is required. We first need to select an appropriate cut-off point that generates false positive and false negative measures. For the selection of the cut-off point, we used the specific SAS Miner *cut-off* node and two strategic selection criteria. The first strategy consisted of finding out the cut-off point that maximizes the number of true positives and negatives, and the second strategy consisted of maximizing the Youden Index (Sensitivity+ Specificity-1). According to the results, both strategies picked the same cut-off point that is 0.49. Thus, the confusion matrix has been created with the cut-off point that can be seen in Table 17.

Table 17 Confusion Matrix

	Prediction=0	Prediction=1
Observation=0	2632	1371
Observation=1	457	3546

The measures that are derived from the confusion matrix were calculated and they can be seen below.

Accuracy: $\frac{TP+TN}{TP+TN+FN+FP}$ = 0.7716

Specificity: $\frac{TN}{TN+FP}$ = 0.6575

Sensitivity: $\frac{TP}{TP+FN}$ = 0.8858

Precision: $\frac{TP}{TP+FP}$ = 0.7211

11. Robustness Checks

Robustness checks are very common when researches want to prove how stable are the coefficient estimates while the regression is modifying, for instance, adding or excluding some estimators. In addition, McFadden (2010) points out that robustness checks are used for testing if there is "wrong-way" causalities in a model. We carried out these checks with the aim of confirming if we created correctly our dependent variable and the derived variables from the paternity reforms (First Permission and Second Permission). In order to check if we accomplished our goals about the statistical analysis, we implemented three models for the robustness check.

11.1 Creating a "Placebo" dependent variable and repeating the final model

We used Permission as the dependent variable for the statistical analysis. It was created from two key questions of the survey. For the first robust analysis, we create an artificial random binary variable ("placebo") that is going to be used as our new dependent variable. In particular, we used a specific function of RStudio that helped us to create a binary variable that was random. The placebo has not any relation with either the real dependent variable or the independent variables that are related with job conditions. We next replace our dependent variable Permission with the new invented binary variable and replicate our final model to evaluate the effects of both paternity reforms and parents' conditions with selected variables. As it was done in the final model, we also interact gender with the rest of the selected variables. We pay special attention to the policy change variables (first permission and second permission) and their interactions with fathers. Most significantly, we want to see whether the real policy variables have a significant statistical effect on the invented dependent variable. Table 18 below presents the p-values for some selected parameters of the model. The detailed output of the regression model can be seen in Appendix F.

Table 18 p-Values of the new model with the Placebo dependent variable

		Pr(>IzI)		Pr(>IzI)
(Intercept)		0.7499	**SITU(Professional Situation):**	
GENDER:			*Wage Earner Public Sector*	0.2026
MALE		0.6635	**TRAREMCONY(Employed Parner):**	

FIRSTPERMISSION:			*NO*	0.9054
YES		0.3994	**NHIJOSLT5(Number of children)**	0.2992
SECOND PERMISSION:			**DCOM(Months in the job)**	0.001
YES		0.2734	**EDAD5 (AGE):**	
MALE:FIRSTPERMISSION.YES		0.5115	*EDAD25-49*	0.2826
MALE:SECONDPERMISSION.YES		0.5144	*EDAD50-65*	0.0765

We observe that none of the policy change variables and their interactions are significant at standard significance levels. A deeper analysis of the full model for the new target variable reveals that almost any of the independent variables are significant except for the variable DCOM (months in the job) for the new target variable (see Appendix F). This was the expected outcome given that the target variable is invented. However, in the original final model, there were several statistically significant relations between the independent variables and the target variable Permission. Thus, we can confirm that our dependent variable is correctly structured, as the results of the model with the placebo are not significant.

11.2 Creating two Placebo variables for the policy changes and repeating the final model

The next robustness check focuses on the paternity permission variables: First Permission and Second Permission. In order to check whether these variables are well-structured, we generated two placebo variables, one for each policy change. The first permission was created as a binary variable that gives value 0 to the observations (individuals) that were interviewed before the first paternity reform in March 2007 and gives value 1 to the individuals that were interviewed after that date. The variable was created for capturing the casual effect of the first reform on the take-up rates of the fathers. The second permission was also created as a binary variable, and it gives the value 1 to the observations that correspond to individuals that were interviewed after the second reform (January 2017), and zero otherwise. The variable was created for capturing the casual effect of the second reform on the take-up rates of the fathers.

For creating placebo variables for these policies, we reproduce the two policy variables by changing the dates of the reforms. We assigned an invented policy date before that of March 2007 to the placebo of the first permission. We follow the same strategy to create a placebo for the second permission reform. We recreated a new variable by changing the date of the paternity reform that brought 4 weeks leave.

On behalf of carrying out the robustness check for the first permission variable, we first created an artificial first permission variable by assigning a fake date (January 2006) to the first paternity reform. Secondly, we reproduced the final model with the real dependent variable, all the selected variables (except the variable of second permission) and the corresponding interactions. Lastly, we divided the dataset into two parts: observations before March 2007 (the real reform date) and after March 2007. The model is estimated for the observations before March 2007.We wanted to check whether the placebo policy variable could have a significant statistical effect in our regression model even though there was no such a real paternity policy reform at that date.

Table 19 The model with the Placebo first policy variable

	Pr(>IzI)		Pr(>IzI)
(Intercept)	0.0261	**CCAA:**	
GENDER:		*NOT ENCOURIGING REGIONS*	0.5142
MALE	4.36e-05	**DCOM(Months in the job):**	1.33e-08
PLACEBOFIRSTPERMISSION:		**EDAD5 (AGE):**	
YES	0.4950	*EDAD25-49*	0.4657
MALE:PLACEBOFIRSTPERMISSION.YES	0.9941	*EDAD50-65*	0.9884
NHIJOSLT5(Number of children)	0.0341	**MALE:EMPLOYEDPARTNER**	2.05e-06
NFORMA(Level of education)		**MALE:TYPEOFCONTRACT**	0.0426
LOWER EDUCATION	6.03e-11	**MALE:DCOM**	0.4566

Table 19 shows the results for a selected group of variables. The focus is on the p-values of the artificial policy variable and its interaction. Neither the invented first permission variable nor its interaction with fathers are significant. This makes sense since there was not a paternity permission reform in those years, therefore the placebo should not be significant, especially the interaction with the Male character (the fathers). The complete output of the regression model can be seen in Appendix G.

For the robustness check of the second permission variable, we first invented a new variable by assigning a fake date (January 2014) for the second paternity reform. The real date of the reform was January 2017. Secondly, we divided the data set again, and we kept the observations from March 2007 to December 2016, that is the period between the first and the second reform. The idea was to consider a period where the second permission does not exist and without the interference of any other reform. Lastly, we estimated the regression model with the placebo variable and the rest of the selected variables of the model (except for the first permission), interactions also included. We wanted to assess whether the placebo variable could be statistically significant even if the date for the permission reform was not accurate.

Table 20 The model with the Placebo second policy variable

	Pr(>IzI)		Pr(>IzI)
(Intercept)	0.7166	**CCAA:**	
GENDER:		*NOT ENCOURIGING REGIONS*	0.1694
MALE	4.96e-15	**DCOM(Months in the job):**	1.67e-15
PLACEBOSECONDPERMISSION:		**EDAD5 (AGE):**	
YES	0.2893	*EDAD25-49*	0.0112
MALE:PLACEBOSECONDPERMISSION.YES	0.2884	*EDAD50-65*	0.1011
NHIJOSLT5(Number of children)	<2e-16	**MALE:EMPLOYEDPARTNER**	<2e-16
NFORMA(Level of education)		**MALE:TYPEOFCONTRACT**	0.2525
LOWER EDUCATION	0.0004	**MALE:DCOM**	0.7243

Table 20 presents the results for some variables of the model. We want to emphasize that the placebo variable of the second policy and its interaction with the fathers are not statistically significant, as we expected.

To sum up, while the models with the invented policy variables do not provide significant effects of these variables, both real policy variables were significant in the regression analysis of the previous section. Based on these results, we can assume that our real policy variables (first permission and second permission) are well-structured and accurate.

12. Conclusions, Limitations and Future Work

Over the last years, the duration of the paternity leave has been growing to give more rights to fathers on childcare activities and to bring gender equality to the family life. In Spain, the last extension brought 8 weeks of permission and entered into force the 1st of April 2019. It is expected to extend the current paternity leave to 16 weeks in 2021. This project aims at studying the effect of the paternity permission on the decision of taking a childbirth leave, and analyzing in a comprehensive way the factors that influence the take-up rate of the paternity leave To evaluate the effects of the reforms of the paternity leave, we considered one target group (the fathers) and one control group (the mothers) to examine differences between the fathers and the mothers.

We created a database from the EAPS survey of the INE with 46 periods of quarterly microdata. The database comprised the years between 2005 and 2018. This survey includes valuable information from different aspects. For instance, demographic factors, individuals´ professional and academic situation, and employed/ unemployed conditions. To prepare the database, we first gathered the individuals who have children less than 5 years old, as the aim of the project is to analyze the take-up rate of the parental permissions after a childbirth. Among parents, we only selected wage earners as it was necessary to keep the individuals who were registered in the Social Security System and who were in possession of a parental permission.

Results from the benchmark model show that the first and the second reforms of the paternity leave have a positive effect on the usage rate of the parental leave among fathers. The policies are not statistically significant when they do not have an interaction with the Male character of the gender variable.

With respect to the estimation of the final model, our study highlights that some characteristics of the parents are statistically significant on the decision of taking a parental permission. The determinants that have a positive impact on the take-up rate of parental leave among men and women are: to have an indefinite contract, to work in the public sector, to have a full-time working contract, to live in a region that is considered one of the encouraging regions, to possess a high education level, to have a partner who is working in either the private sector or the family business, and to have a high number of children who are less than 5 years old. The gender variable is statistically significant as the take-up rate of parental leave for the mothers is higher than for the fathers. Therefore, being a mother has a positive effect on taking a parental leave. The variables derived from the policy change (first permission and second permission) are affecting positively as well. The determinants that affect negatively the take-up rates of both genders are: to have a professional occupation as military or director, the time (monthly) that the individual has been in the same job and to have a partner from the group that consists undefined job situations and unemployment.

We observed some differences on the estimation when considering the interactions with the fathers. The determinants that affect positively the fathers on their use of the parental leave are: to work in the public sector, to be in the period of the first paternity policy, to be in the period of the second paternity policy, and to work in a professional activity such as Food, Textile, Extractive Industries or Transportation sectors. The determinants that are affecting negatively the take-up rate of the paternity leave are: to have an indefinite contract, the number of children who are less than 5 years old, and to have a partner from the group that consists undefined job situations and unemployment. Although the indefinite contract and number of children have positive significantly effect among men and women, their interactions with fathers have negative effect on the dependent variable.

We faced some limitations while doing the study. First of all, it does not exist an official survey that makes a record of the parental leave. Instead, we used the EAPS. The EAPS contains some variables that can be related with the parental leave, and it also has relevant information on parents' characteristics. Secondly, due to the survey design, it was not possible to distinguish between different types of parental leave. In other words, if an individual was not working due to the birth of a child in the reference week, the only response possible would be "permission for a childbirth". But this response embraces different possibilities for a father. The first option is that he can use his own paternity leave, the second option is that he can use the maternity leave that was transferred from the mother, and the last option is that he can use an unpaid leave, which is not a parental leave. Thus, it is not completely clear whether fathers are using the paternity permission or other type of leave that can be used for childcare. Despite of this, the general response has sufficient perspective for creating a dependent variable. Lastly, the distribution of the dependent variable is unbalanced, as there were very few individuals who had a permission compared to the ones who did not have it. In the survey questions, the reference week refers to one week before the interview day. Thus, the use of a parental permission before the reference week could not be detected.

Despite these limitations, we obtained some results that give us a better understanding of the factors affecting parents' decision of taking a leave after the birth of a child. For future work, following Escot et al. (2013), we could include in the model a macroeconomic variable such as the unemployment rate to capture the effect of the economic crisis. Moreover, as we have further data, the last paternity leave reform should be included in the model to analyze whether successive extensions of the paternity leave have a positive impact on its usage.

13.Reference

Akgunduz, Yusuf E., & Plantenga J. (2012). 'Labour market effects of parental leave in Europe',Cambridge Journal of Economics, Vol: 37, No.4 pp. 845-862.

Angrist, J. D. & Pischke, J.-S. (2009): Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press

Belle J. (2016). Paternity and parental leave policies across the European Union. Retrieved from https://www.rand.org/pubs/research_reports/RR1666.html

Borraz, M., Sánchez, R., & Ordaz. A. (2019, June 16). El número de países con permisos de paternidad aumenta casi un 40% en los últimos nueve años [The number of countries with paternity leave increases by almost 40% in the last nine years]. El Diario. Retrieved from https://www.eldiario.es/sociedad/

Columbia University Mailman School. (n.d.) Difference-in-Difference-Estimation. Retrieved from https://www.mailman.columbia.edu/

Escot, L., Fernández-Cornejo, J. A. and Poza, C. (2013). 'Fathers' use of childbirth leave in Spain – The effects of the 13-day paternity leave', Population Research and Policy Review, Vol. 33, No. 3, pp. 419–453.

European Platform for Investing in Children [EPIC]. (2018). Paternity and parental leave policies across the European Union. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/a8464ad8-9abf-11e8-a408-01aa75ed71a1/language-en

McFadden D. (1999). Robust Methods in Econometrics. Retrieved from https://eml.berkeley.edu/~mcfadden/e240b_f01/ch7.pdf

Meil G., Rogero J., & Romero P. (2017, February). Parental leave as a tool for gender equality. La Caixa Foundation. Retrieved from https://observatoriosociallacaixa.org/en/

Miani, C. & S. Hoorens. 2014. 'Single parents and employment in Europe.' European Commission. Retrieved from http://ec.europa.eu/justice/gender-equality/files/documents/140502 _gender_equality_workforce_ssr3_en.pdf

Pacific Community. (2019). Toolkit Training and DDI Production Workshop [PowerPoint slides]. Retrieved from http://www.spc.int/DigitalLibrary/Doc/SDD/Workshops/2019/South_South_Capacity_Workshop/South_South_Capacity_NC_19_Microdata_Documentation_Production.pptx

Reich, N. (2011). 'Predictors of Fathers' Use of Parental Leave in Germany ', Population Review, Vol. 50, No. 2, pp. 1–22.

Rouse, M. (2019). Logistic Regression. Retrieved from https://searchbusinessanalytics.techtarget.com/definition/logistic-regression

Sayad, S. (2019). Logistic Regression. Retrieved from https://www.saedsayad.com/

Seguridad Social. (n.d.). General Scheme. Retrieved from http://www.seg-social.es/wps/portal/wss/internet/Trabajadores?1dmy&urile=wcm%3apath%3a%2FPOIN_Contenidos_en%2FInternet%2F4983%2F10935%2F51113%2F51120%2F#51169

Spanish Statistical Office [INE]. (2019). Economically Active Population Survey. Retrieved from https://www.ine.es/en/welcome.shtml

United Nations Children's Fund [UNICEF]. (2017). Gender Equality: Glossary of Terms and Concepts. Retrieved from https://www.unicef.org/rosa/media/1761/file/Gender%20glossary%20of%20terms%20and%20concepts%20.pdf

WORLD Policy Analysis Center. (2018). Paid Parental Leave: A Detailed Look at Approaches across OECD Countries. Retrieved from https://www.worldpolicycenter.org/sites/default/files/WORLD%20Report%20-%20Parental%20Leave%20OECD%20Country%20Approaches_0.pdf

Zhelyazkova, N. (2013). 'Fathers' use of parental leave. What do we know? 'Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT)

Appendix A

```
#===============================================#
#   File: SelectSample.R
#_______________________________________________#
selectSample <- function(fichero_micro,the_dir){
####################     INICIAL     #########################
  start.time <- Sys.time()
  cat("\n")
  cat("\n Inicio: ")
  print.Date(start.time)
  t0 <- proc.time()
###########################################################
if(!require(dplyr)){install.packages("dplyr")}
if(!require(data.table)){install.packages("data.table")}
library(dplyr)
library(data.table)
name_fichero <- substr(fichero_micro,5,10)
load(fichero_micro,.GlobalEnv)
datoslt5 <-fichero_salida%>%
  dplyr::arrange(NVIVI)%>%
  dplyr::filter(EDAD5=="00" & (NPADRE!="00" & NMADRE!="00"))
distinct_nvivi <- datoslt5%>% distinct(NVIVI)
datos <- inner_join(fichero_salida,distinct_nvivi,by="NVIVI")
npadretot <- datoslt5%>%
  group_by(NVIVI,NPADRE)%>%
  summarise(NPADRETOT= n())
datosAF <- left_join(datos,npadretot,by=c("NVIVI","NPERS"="NPADRE"))
nmadretot <- datoslt5%>%
  group_by(NVIVI,NMADRE) %>%
  summarise(NMADRETOT= n())
```

```
datosAFM <- left_join(datosAF,nmadretot,by=c("NVIVI","NPERS"="NMADRE"))
datosNHLT5 <- datosAFM%>%
  dplyr::mutate(NHIJOSLT5=NPADRETOT) %>%
  dplyr::mutate(NHIJOSLT5=ifelse(is.na(NHIJOSLT5),NMADRETOT,NHIJOSLT5)) %>%
  dplyr::select(-c(NPADRETOT,NMADRETOT))
datosNRKIDS1 <- datosNHLT5%>%
  mutate(NOPAR=ifelse(NCONY=="00" & !is.na(NHIJOSLT5),1,0))
addID2 <- datosNRKIDS1%>%
  group_by(NVIVI)%>%
  summarise(idNOPAR=sum(NOPAR,na.rm=TRUE))
datosNRKIDS2 <- inner_join(datosNRKIDS1,addID2,by="NVIVI")
datosNRKIDS <- datosNRKIDS2%>%
  filter(idNOPAR==0)%>%
  select(-c(idNOPAR,NOPAR))
datosIDCONY <- datosNRKIDS%>%
  filter(NCONY!="00")%>%
  rowwise() %>%
  mutate(NCONYMIN=min(as.numeric(NPERS),as.numeric(NCONY)))%>%
  ungroup()%>%
  mutate(NEWPAR_NOKIDS=ifelse(!is.na(NCONYMIN) & is.na(NHIJOSLT5),1,0))
datosIDCONY1 <- datosIDCONY%>%
  filter(NEWPAR_NOKIDS==0)%>%
  select(-NEWPAR_NOKIDS)
mydata <- datosIDCONY1%>%
  group_by(NVIVI,NCONYMIN) %>%
  mutate(EDAD5CONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(EDAD5),lag(EDAD5)))%>%
  mutate(NAC1CONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(NAC1),lag(NAC1)))%>%

mutate(NFORMACONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(NFORMA),lag(NFORMA)))
%>%

mutate(TRAREMCONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(TRAREM),lag(TRAREM)))%
>%
```

```
  mutate(OCUP1CONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(OCUP1),lag(OCUP1)))%>%
  mutate(ACT1CONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(ACT1),lag(ACT1)))%>%
  mutate(SITUCONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(SITU),lag(SITU)))%>%
mutate(DUCON1CONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(DUCON1),lag(DUCON1)))
%>%
  mutate(DCOMCONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(DCOM),lag(DCOM)))%>%
mutate(PARCO1CONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(PARCO1),lag(PARCO1)))%>
%
  mutate(AOICONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(AOI),lag(AOI)))%>%
  mutate(CSECONY=ifelse(as.numeric(NPERS)==NCONYMIN,lead(CSE),lag(CSE)))
rm(datoslt5,distinct_nvivi,datos,npadretot,datosAF,nmadretot,datosAFM,datosNHLT5,datosN
RKIDS,datosNRKIDS1,datosNRKIDS2,addID2,datosIDCONY,datosIDCONY1)
mydata <- setDT(mydata)
save(mydata,file=paste0(the_dir, "/", "select",name_fichero, ".RData"))
# Mensaje final ##########################################
end.time <- Sys.time()
cat("\n")
cat("\n Fin del proceso de lectura: ")
print.Date(end.time)
TTotal <- proc.time() - t0
tiempo <- TTotal[3]
if(tiempo < 60){
  cat(paste("\n Tiempo transcurrido:", format(round(tiempo, 2), nsmall = 2), "segundos"))
}else{
  if(tiempo< 3600 & tiempo >= 60){
    cat(paste("\n Tiempo transcurrido:", format(round(tiempo/60, 2), nsmall = 2), "minutos"))
  }else{
    cat(paste("\n Tiempo transcurrido:", format(round(tiempo/3600, 2), nsmall = 2), "horas"))
  }
}
##############################################################
```

```
}
load("mysample.RData")
#wageearners
SOLOASA<-mysample%>%
  #  dplyr::filter(RZNOTB=="02" | RZNOTB=="03" | VINCUL=="02" | RZDIFH=="02" ) %>%
 filter(SITU=="07" | SITU=="08")
#We created a new variable about the parental leave
SOLOASA2_2<-SOLOASA%>%
  mutate(permiso=ifelse((RZNOTB=="02"  |  RZNOTB=="03"  |  VINCUL=="02"  |  RZDIFH=="02"
),1,0))%>%
  filter(AOI=="03" | AOI=="04"| permiso==1)
#CHANGE OF POLICY WE SHOULD PUT 0,1,2,..)
SOLOASA3<-SOLOASA2_2%>%
  mutate(CAMBIOPOLI = case_when(CICLO <= 138 ~ '0SEM',
                          CICLO <= 177  ~ '2SEM',
                          CICLO <= 183 ~ '4SEM',
                          CICLO <= 185 ~ '5SEM',
                          TRUE ~ 'NA'))
#as the last policy do not have sufficient observations we are deleting
SOLOASA3.1<-SOLOASA3%>%
  filter(CAMBIOPOLI != '5SEM' )
SOLOASA3.2<-SOLOASA3.1%>%
  mutate(FIRSTPERMISSION = case_when(CAMBIOPOLI=='0SEM'~'0',
                            CAMBIOPOLI=='2SEM'~'1',
                            CAMBIOPOLI=='4SEM'~'1',
                            TRUE ~ 'NA'))
SOLOASA3.3<-SOLOASA3.2%>%
  mutate(SECONDPERMISSION= case_when(CAMBIOPOLI=='0SEM'~'0',
                            CAMBIOPOLI=='2SEM'~'0',
                            CAMBIOPOLI=='4SEM'~'1',
                            TRUE ~ 'NA'))
#TO MAKE THE UNDERSAMPLE
```

```
library(caret)
set.seed(9560)
SOLOASA3.4$permiso<-factor(SOLOASA3.3$permiso)
down_trainSON <- downSample(x = SOLOASA3.3,
                     y = SOLOASA3.4$permiso)
```

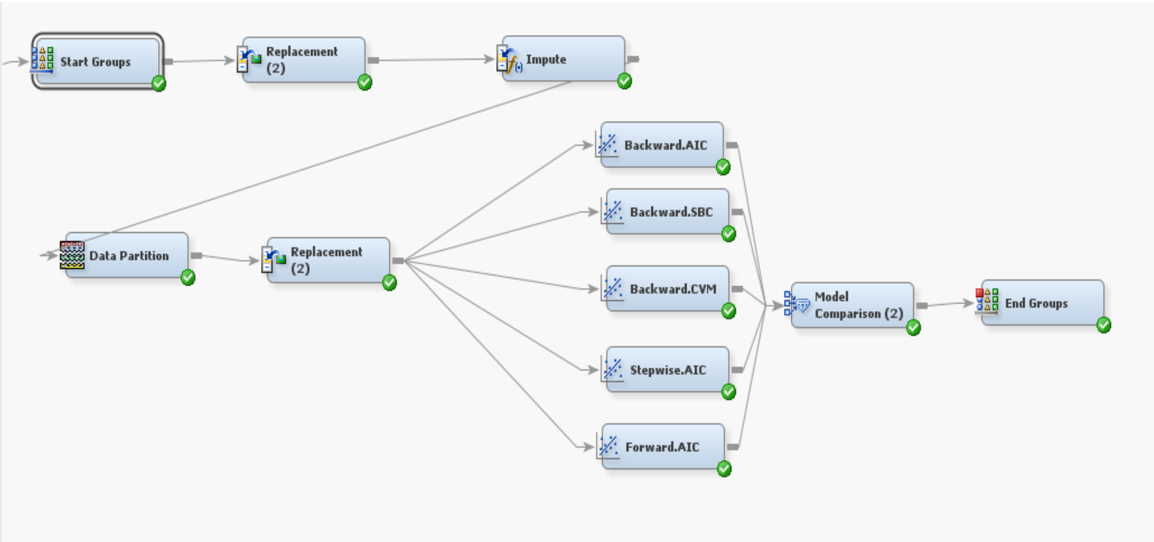
Appendix B

```
frec1 <- down_trainSON %>% group_by(SEXO1) %>%
  summarize(count = n(), sumper=sum(FACTOREL)) %>%
  mutate(pct = sumper/sum(sumper))  # find percent of total
frec1
ggplot(frec1, aes(SEXO1, pct, fill = SEXO1,label = scales::percent(pct))) +
  geom_bar(stat='identity',width = 0.5) +
  geom_col(position = 'dodge') +
  geom_text(position = position_dodge(width = .9),    # move to center of bars
        vjust = -0.5,    # nudge above top of bar
        size = 3) +
  scale_y_continuous(labels = scales::percent_format())
```

Appendix C

Variable Name	Role ▲	Measurement Level
ACT1	Input	Nominal
ACT1CONY	Input	Nominal
AOI	Input	Nominal
AOICONY	Input	Nominal
CAMBIOPOLI	Input	Nominal
CCAA	Input	Nominal
CICLO	Input	Nominal
CSE	Input	Nominal
DCOM	Input	Interval
DUCON1	Input	Nominal
ECIV1	Input	Nominal
EDAD5	Input	Nominal
EDAD5CONY	Input	Nominal
EDADEST	Input	Interval
FIRSTPERMISSION	Input	Binary
NAC1	Input	Nominal
NAC1CONY	Input	Nominal
NFORMA	Input	Nominal
NFORMACONY	Input	Nominal
NHIJOSLT5	Input	Interval
OCUP1	Input	Nominal
OCUP1CONY	Input	Nominal
PARCO1	Input	Binary
PARCO1CONY	Input	Binary
SECONDPERMISSION	Input	Binary
SEXO1	Input	Binary
SITU	Input	Nominal
SITUCONY	Input	Nominal
TRAPLU	Input	Binary
TRAREMCONY	Input	Binary
permiso	Target	Binary

Appendix D



Appendix E

Analysis of Maximum Likelihood Estimates

Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Estimador estandarizado	Exp(Est)
Intercept			1	-1.6106	0.2219	52.70	<.0001		0.200
DUCON1*SEX01	1	1	1	-0.0848	0.0346	6.01	0.0142		0.919
FIRSTPERMISSION*SEX01	0	1	1	-0.2267	0.0383	35.03	<.0001		0.797
GRP_ACT1*SEX01	1	1	1	0.0208	0.0732	0.08	0.7767		1.021
GRP_ACT1*SEX01	2	1	1	0.1123	0.0517	4.72	0.0298		1.119
GRP_ACT1*SEX01	3	1	1	0.0364	0.0699	0.27	0.6025		1.037
GRP_ACT1*SEX01	4	1	1	-0.0190	0.0492	0.15	0.6995		0.981
GRP_OCUP1*SEX01	1	1	1	-0.00067	0.0551	0.00	0.9903		0.999
GRP_OCUP1*SEX01	2	1	1	0.0198	0.0476	0.17	0.6768		1.020
GRP_OCUP1*SEX01	3	1	1	0.0247	0.0612	0.16	0.6870		1.025
GRP_OCUP1*SEX01	4	1	1	-0.1034	0.0781	1.75	0.1857		0.902
GRP_REP_SITUCONY*SEX01	1	1	1	-0.6214	0.0555	125.37	<.0001		0.537
GRP_REP_SITUCONY*SEX01	2	1	1	-0.00044	0.0530	0.00	0.9933		1.000
GRP_REP_SITUCONY*SEX01	3	1	1	0.0369	0.0400	0.85	0.3556		1.038
IMP_REP_DCOM*SEX01	1		1	0.000274	0.000380	0.52	0.4698		1.000
NAC1*SEX01	1	1	1	-0.0800	0.0713	1.26	0.2617		0.923
NAC1*SEX01	2	1	1	0.2064	0.1250	2.73	0.0988		1.229
NHIJOSLT5*SEX01	1		1	-0.1304	0.0487	7.16	0.0075		0.878
PARC01*SEX01	1	1	1	-0.0465	0.0726	0.41	0.5217		0.955
REP_CCAA*SEX01	ENCOURAGING	1	1	0.0210	0.0250	0.71	0.4000		1.021
REP_EDAD5*SEX01	16-24	1	1	-0.2546	0.2042	1.55	0.2126		0.775
REP_EDAD5*SEX01	25-49	1	1	-0.0996	0.1442	0.48	0.4897		0.905
REP_NFORMA*SEX01	HIGH EDUCATION	1	1	-0.0219	0.0287	0.59	0.4441		0.978
REP_NFORMACONY*SEX01	HIGH EDUCATION	1	1	0.00198	0.0271	0.01	0.9417		1.002
SECONDPERMISSION*SEX01	0	1	1	-0.1578	0.0347	20.70	<.0001		0.854
SEX01*SITU	1	07	1	0.2080	0.0512	16.48	<.0001		1.231
SEX01*TRAREMCONY	1	1	1	-0.6931	0.0358	373.98	<.0001		0.500

DUCON1	1	1	0.2785	0.0346	64.76	<.0001		1.321
FIRSTPERMISSION	0	1	-0.1790	0.0383	21.83	<.0001		0.836
GRP_ACT1	1	1	-0.0795	0.0732	1.18	0.2776		0.924
GRP_ACT1	2	1	0.0802	0.0517	2.41	0.1205		1.084
GRP_ACT1	3	1	-0.0218	0.0699	0.10	0.7548		0.978
GRP_ACT1	4	1	0.0791	0.0492	2.58	0.1081		1.082
GRP_OCUP1	1	1	0.0163	0.0553	0.09	0.7686		1.016
GRP_OCUP1	2	1	0.0837	0.0476	3.09	0.0788		1.087
GRP_OCUP1	3	1	-0.3128	0.0612	26.10	<.0001		0.731
GRP_OCUP1	4	1	0.0162	0.0782	0.04	0.8362		1.016
GRP_REP_SITUCONY	1	1	-1.0845	0.0555	381.40	<.0001		0.338
GRP_REP_SITUCONY	2	1	0.1007	0.0530	3.61	0.0576		1.106
GRP_REP_SITUCONY	3	1	0.2362	0.0400	34.95	<.0001		1.266
IMP_REP_DCOM		1	-0.00324	0.000380	72.73	<.0001	-0.1188	0.997
IMP_REP_EDADEST		1	-0.0110	0.00506	4.76	0.0291	-0.00287	0.989
NAC1	1	1	0.0316	0.0713	0.20	0.6582		1.032
NAC1	2	1	0.2084	0.1250	2.78	0.0956		1.232
NHIJOSLT5		1	1.0538	0.0488	467.11	<.0001	2.9165	2.869
PARC01	1	1	0.3539	0.0726	23.79	<.0001		1.425
REP_CCAA	ENCOURAGING	1	0.0505	0.0250	4.09	0.0432		1.052
REP_EDAD5	16-24	1	0.1456	0.2043	0.51	0.4760		1.157
REP_EDAD5	25-49	1	-0.1022	0.1442	0.50	0.4783		0.903
REP_NFORMA	HIGH EDUCATION	1	0.0668	0.0317	4.44	0.0352		1.069
REP_NFORMACONY	HIGH EDUCATION	1	0.0492	0.0272	3.27	0.0704		1.050
SECONDPERMISSION	0	1	-0.1953	0.0347	31.63	<.0001		0.823
SEX01	1	1	-0.9981	0.1936	26.57	<.0001		0.369
SITU	07	1	0.2458	0.0512	23.01	<.0001		1.279
TRAREMCONY	1	1	-0.9808	0.0359	748.49	<.0001		0.375

Appendix F

```
                                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                                  7.854e-02  2.464e-01   0.319  0.74991
SEXO1 == 1TRUE                              -1.963e-01  4.512e-01  -0.435  0.66354
FIRSTPERMISSION1                             4.381e-02  5.199e-02   0.843  0.39947
SECONDPERMISSION1                           -6.674e-02  6.093e-02  -1.095  0.27340
GRP_REP_SITUCONY                             1.442e-02  2.682e-02   0.538  0.59090
SITU08                                       6.677e-02  5.241e-02   1.274  0.20267
TRAREMCONY6                                  6.420e-03  5.402e-02   0.119  0.90540
GRP_ACT1                                    -6.436e-03  2.012e-02  -0.320  0.74904
DUCON16                                     -1.099e-06  5.231e-02   0.000  0.99998
IMP_REP_DCOM                                 1.081e-03  3.369e-04   3.208  0.00134 **
IMP_REP_EDADEST                             -2.650e-03  5.282e-03  -0.502  0.61589
NAC12                                        1.479e-01  1.503e-01   0.984  0.32515
NAC13                                       -2.560e-02  8.149e-02  -0.314  0.75338
NHIJOSLT5                                   -3.899e-02  3.757e-02  -1.038  0.29926
GRP_OCUP1                                   -2.328e-02  2.357e-02  -0.988  0.32329
PARCO16                                      6.695e-02  4.512e-02   1.484  0.13789
REP_NFORMALOWER EDUCATION                    4.049e-02  5.347e-02   0.757  0.44890
REP_NFORMACONYLOWER EDUCATION                9.019e-03  4.023e-02   0.224  0.82260
REP_EDAD525-49                              -1.536e-01  1.430e-01  -1.074  0.28261
REP_EDAD550-65                               1.399e+00  7.902e-01   1.771  0.07658 .
REP_CCAANOT                                 -1.300e-02  3.982e-02  -0.326  0.74414
SEXO1 == 1TRUE:FIRSTPERMISSION1             -5.692e-02  8.671e-02  -0.656  0.51155
SEXO1 == 1TRUE:SECONDPERMISSION1            -6.734e-02  1.033e-01  -0.652  0.51443
SEXO1 == 1TRUE:GRP_REP_SITUCONY             -1.223e-02  4.269e-02  -0.286  0.77461
SEXO1 == 1TRUE:SITU08                       -1.084e-02  9.681e-02  -0.112  0.91086
SEXO1 == 1TRUE:TRAREMCONY6                  -1.485e-02  8.543e-02  -0.174  0.86200
SEXO1 == 1TRUE:GRP_ACT1                      4.712e-02  3.107e-02   1.516  0.12945
SEXO1 == 1TRUE:DUCON16                      -7.865e-02  8.958e-02  -0.878  0.37992
SEXO1 == 1TRUE:IMP_REP_DCOM                 -6.555e-04  5.090e-04  -1.288  0.19779
SEXO1 == 1TRUE:IMP_REP_EDADEST               2.649e-03  8.052e-03   0.329  0.74216
SEXO1 == 1TRUE:NAC12                        -2.140e-01  2.687e-01  -0.796  0.42579
SEXO1 == 1TRUE:NAC13                        -1.135e-02  1.237e-01  -0.092  0.92689
SEXO1 == 1TRUE:NHIJOSLT5                     1.586e-02  7.010e-02   0.226  0.82100
SEXO1 == 1TRUE:GRP_OCUP1                     4.465e-02  3.181e-02   1.404  0.16035
SEXO1 == 1TRUE:PARCO16                      -1.378e-02  1.508e-01  -0.091  0.92721
SEXO1 == 1TRUE:REP_NFORMALOWER EDUCATION    -2.439e-02  8.844e-02  -0.276  0.78273

SEXO1 == 1TRUE:REP_NFORMACONYLOWER EDUCATION -6.238e-02  7.059e-02  -0.884  0.37684
SEXO1 == 1TRUE:REP_EDAD525-49                9.390e-02  3.139e-01   0.299  0.76481
SEXO1 == 1TRUE:REP_EDAD550-65               -1.335e+00  8.665e-01  -1.540  0.12352
SEXO1 == 1TRUE:REP_CCAANOT                  -4.095e-02  6.814e-02  -0.601  0.54782
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

Appendix G

```
Coefficients:
                                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                                  0.1130391  0.3114201   0.363 0.716620
SEXO1 == 1TRUE                              -5.7942863  0.7402100  -7.828 4.96e-15 ***
PLACEBOSECONDPER1                           -0.0699861  0.0660585  -1.059 0.289392
GRP_REP_SITUCONY                             0.1833364  0.0352496   5.201 1.98e-07 ***
SITU08                                      -0.0024777  0.0687243  -0.036 0.971241
TRAREMCONY6                                  0.4474424  0.0730116   6.128 8.88e-10 ***
GRP_ACT1                                     0.0807716  0.0254795   3.170 0.001524 **
DUCON16                                     -0.7199924  0.0651119 -11.058  < 2e-16 ***
IMP_REP_DCOM                                -0.0034133  0.0004286  -7.963 1.67e-15 ***
IMP_REP_EDADEST                             -0.0182712  0.0067207  -2.719 0.006555 **
NAC12                                       -0.3465302  0.1799105  -1.926 0.054089 .
NAC13                                       -0.3386787  0.0983582  -3.443 0.000575 ***
NHIJOSLT5                                    1.2620527  0.0596043  21.174  < 2e-16 ***
GRP_OCUP1                                   -0.0362174  0.0295772  -1.225 0.220762
PARCO16                                     -0.8807893  0.0555279 -15.862  < 2e-16 ***
REP_NFORMALOWER EDUCATION                   -0.2365648  0.0674513  -3.507 0.000453 ***
REP_NFORMACONYLOWER EDUCATION               -0.1570335  0.0515390  -3.047 0.002312 **
REP_EDAD525-49                              -0.4490136  0.1772164  -2.534 0.011286 *
REP_EDAD550-65                              -1.2176853  0.7428506  -1.639 0.101170
REP_CCAANOT                                 -0.0702983  0.0511662  -1.374 0.169466
SEXO1 == 1TRUE:PLACEBOSECONDPER1             0.1574263  0.1482885   1.062 0.288407
SEXO1 == 1TRUE:GRP_REP_SITUCONY              0.7789730  0.0627645  12.411  < 2e-16 ***
SEXO1 == 1TRUE:SITU08                       -0.9201472  0.1604099  -5.736 9.68e-09 ***
SEXO1 == 1TRUE:TRAREMCONY6                   2.8859318  0.1566950  18.418  < 2e-16 ***
SEXO1 == 1TRUE:GRP_ACT1                     -0.1327879  0.0508675  -2.610 0.009042 **
SEXO1 == 1TRUE:DUCON16                       0.1775927  0.1585162   1.120 0.262567
SEXO1 == 1TRUE:IMP_REP_DCOM                  0.0003026  0.0008581   0.353 0.724397
SEXO1 == 1TRUE:IMP_REP_EDADEST               0.0245657  0.0130297   1.885 0.059381 .
SEXO1 == 1TRUE:NAC12                         0.8882072  0.4299177   2.066 0.038829 *
SEXO1 == 1TRUE:NAC13                        -0.0868035  0.2231785  -0.389 0.697319
SEXO1 == 1TRUE:NHIJOSLT5                    -0.3652612  0.1094595  -3.337 0.000847 ***
SEXO1 == 1TRUE:GRP_OCUP1                     0.0833823  0.0497872   1.675 0.093979 .
SEXO1 == 1TRUE:PARCO16                       0.3925928  0.3371397   1.164 0.244229
SEXO1 == 1TRUE:REP_NFORMALOWER EDUCATION     0.1431772  0.1436899   0.996 0.319040
```

```
SEXO1 == 1TRUE:REP_NFORMACONYLOWER EDUCATION -0.0938519  0.1188444  -0.790 0.429701
SEXO1 == 1TRUE:REP_EDAD525-49                 0.1046137  0.5150815   0.203 0.839056
SEXO1 == 1TRUE:REP_EDAD550-65                 1.3735512  0.9655297   1.423 0.154856
SEXO1 == 1TRUE:REP_CCAANOT                   -0.1346292  0.1111586  -1.211 0.225840
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 20240  on 14599  degrees of freedom
Residual deviance: 14113  on 14562  degrees of freedom
AIC: 14189

Number of Fisher Scoring iterations: 6

>
```