



Facultad de Ciencias Económicas y Empresariales

TRABAJO FIN DE GRADO

GRADO EN

Administración y Dirección de Empresas Internacional

ESTRUCTURA SALARIAL ESPAÑOLA

Andrea San Martín Chaler

DIRECTOR

Ignacio García Lautre

CODIRECTOR

Blanca Palacios Navarro

Pamplona-Iruña

11 de Junio del 2014

SUMMARY:

The following report consisting of the description and analysis of the variables which affect the wage structure is based on the salaries that the employees of educational sector perceive. This is studied inside the Spanish context.

The problem of this report was the fact of analyzing the different variables that affect the salary structure because of the effect of discrimination they could make. This report will help people to understand when a variable is significant or not in order to explain the structure of the Spanish salary.

In order to analyze the significance of each variable it is estimated an econometric model based on the data recollected previously from the survey made by INEM. There are analyzed among the report two kinds of variables: Qualitative, such as age or level of study and quantitative such as Base salary or ages worked.

Having done all the study, we conclude with one final econometric model which eliminates the variables which are not relevant and only include the significant ones.

***KEYWORDS**

Mean: is used to refer to one measure of the central tendency either of a probability distribution or of the random variable characterized by that distribution.

Median: The median is the numerical value separating the higher half of a data sample, a population, or a probability distribution, from the lower half.

Coefficient of Variation: is a normalized measure of dispersion of a probability distribution or frequency distribution. It is defined as the ratio of the standard deviation to the mean.

Kurtosis: is any measure of the "peakedness" of the probability distribution of a real-valued random variable.

Skewness: is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

Heteroscedasticity: a collection of random variables is heteroscedastic if there are sub-populations that have different variabilities from others.

Homoscedasticity: a sequence or a vector of random variables is homoscedastic if all random variables in the sequence or vector have the same finite variance.

Multicollinearity: is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy.

INDEX.

1. INTRODUCTION	Page1
2. DESCRIPTIVE AND INFERENTIAL ANALYSIS	Page4
2.1 Descriptive analysis of qualitative variables	Page4
2.1.1 Sex	Page4
2.1.2 Age	Page 4
2.1.3 Control	Page 5
2.1.4 Type of contract	Page 6
2.1.5. Nationality	Page 7
2.1.6. Type of Study	Page 8
2.2 Descriptive analysis of quantitative variables	Page 8
2.2.1 Base Salary	Page 8
2.2.2 Years worked	Page 10
2.2.3 Salary's complements	Page 12
2.3 Relationship of qualitative variables	Page 14
2.3.1 Sex and type of study	Page 14
2.3.2 Age and type of contract	Page 14
2.3.3 Nationality and type of control	Page 15
2.4 Relationship between quantitative variables	Page 16
2.4.1 Base salary and Salary's complements	Page 16
2.4.2 Base salary and years worked	Page 16
2.4.3 Years worked and Salary's complements.	Page 17
2.5 Inferial Analysis	Page 17

2.5.1 1st Contrast of Hypothesis	Page 17
2.5.2 2nd Contrast of Hypothesis	Page 18
2.5.3 3rd Contrast of Hypothesis	Page 19
3. MODEL'S ESTIMATION	Page 20
3.1 Dummies	Page 20
3.2 Heteroscedasticity	Page 22
3.3 ANOVA contrast	Page 23
3.4 Interpretation of the econometric variables	Page 24
3.5 Individual Significance	Page 25
3.6 Study of multicollinearity's problem	Page 26
3.7 Final econometric model	Page 27
4. CONCLUSIONS	Page 28
5. REFERENCES	Page 30
6. ANNEXES	Page 31

1. INTRODUCTION

From the beginning of the humanity, the salary is one of the topics most complicated at the moment of speaking about them due to his social and human implications. A constant study has been carried out on a more equitable relationship between the capital and work in order to improve the standard of living of the employees.

Etymologically speaking, salary is derived from Latin “salarium” which means 'payment of salt'. This term comes from the ancient Roman Empire, where the soldiers payments were made with salt, which was worth its weight in gold, as in ancient times salt was one of the few ways they had of preserving meat.

If we review the history of salary, we note that this concept has suffered several changes over the last few centuries. After its creation in the former Rome, the salary was manifested in slavery. Despite the fact that the slaves did not perceive any salary, his owner had to feed them in order to maintain them in good physical conditions. Subsequently, the slavery finished and the concept of serfdom was extended in medieval times. The vassal had to look after the land of the feudal lord. In return for this, the lord allowed to live the vassal and his family on the land and cultivate their own plots. When both industrial revolutions finished, techniques in salary's management suffered transformations in the incentives, simpler and more humane standards in setting a fixed salary and other innovations which have the objective of achieve a balance between value and labor.

Nowadays the wage structure in Spain is a determining factor in many issues related to the distribution of income and well-functioning labor market. The salary in Spain is regulated in the consolidated text of the Workers' Statute, Articles 26 et seq.

The salary structure is the sum of base salary and salary complements while in some cases there are not any complements. The salary complements are the bonuses that are agreed by collective agreement and may be of different types from which we highlight: personal (capabilities of the worker), the qualification (work-related or jobs-related) and finally, profits and investments (linked to company's performance). There are also bonuses which are compulsory and extra-wage provisions, those that are not taken into account when we have to pay contributions but they have tax withheld at source.

The wage payment is one of the aspects of working conditions that most directly affect the daily lives of workers. Since its early years, the center of the action the “International Labor

Organization” has revolved around the level of wages constantly struggling to establish standards that guarantee and protect the right of workers to earn a fair wage.

Pay equality refers to the concept that individuals doing similar work (or work with the same productivity) should receive the same remuneration, regardless of gender, race, sexual orientation, nationality, religion or any other category.

However despite the theories focused on competitive model in which there is a single average wage and salary structure stability, there is a wide range of inequalities in some aspects such as gender, age, occupation and region where companies operate.

This study is mainly based on the inequalities that exist in Spain. According to the report published by the INE in 2010 the main differences correspond to the geographical distribution as Catalonia, the Basque Country and Madrid have higher wages with respect to the Canary Islands, Extremadura and Galicia, where they provide the lowest wages. On the other hand, looking at the main differences in the activity which plays the employee, we can emphasize that directors and managers receive higher wages, second would be university graduates and finally the wages of workers without qualifications tends to be below the average. The last wage gap corresponds to the distribution of wages by sex, men’s earning has always been higher than that of women.

Especially this work is analyzed in the field of education. Education is a sector which has undergone many changes and cutbacks due to crisis we are experiencing so it is interesting to know more about how this economic situation has affected the distribution of wages and which wage differences have arisen in the public and private sectors.

These wage differences have remained historically over time so today, it is a fact that has not yet reached solved.

In this study, we can find at least three reasons that may explain the existence of these inequalities in today's economy: In first place the differences in jobs, secondly the heterogeneity of workers (non-wage attributes are different, require different types of skills and they vary in terms of the effectiveness of the payment of efficiency wages to increase productivity) and finally the characteristics of firms and labor markets, as well as imperfections that occur in these markets. We can note that in today’s economies the conditions of perfect competition are not met and there are many problems of information and transparency of the different labor markets.

At the same time, this study has been approached from a social dimension given that the wage is determined not only by market forces but also by working collective bargaining.

The main objective of this research is to examine the distribution of wages based on a variety of variables such as sex, occupation, branch of activity, age, type of employment contract, holidays and overtime or company size and show that wage inequalities are grounded in the reasons we have explained previously. The purpose of this study is to investigate wage inequality in the Spanish market in order to contribute to get a better understanding of the phenomenon. Finally, the ultimate aim of it consists of granting a possible solution based on the development of an econometric model that concludes with the problem encountered.

To perform this analysis the necessity of using a database containing large business variables arise. For this, we turn to the Wage Structure Survey done by the INE in order to obtain comparable results on the structure and distribution of salaries among its member States. This Wage Structure Survey is a statistical four-yearly operation, carried out under the EU with a common methodology and criteria contains. This survey investigates the distribution of wages based on a variety of variables such as sex, occupation, branch of activity, age, or size of the company.

In the next section the descriptive and inferential analysis takes place.

2. DESCRIPTIVE AND INFERENTIAL ANALYSIS

2.1 Descriptive analysis of qualitative variables

2.1.1 *Sex:*

The variable Sex has been chosen because nowadays in the education's field there is a wage gap with respect to the wages paid to both sexes. This variable is divided between male and female. The masculine gender is designated by the number 1 and the female gender is designated by the number 6.

Frequency distribution of SEX, observations 1-6747

Frequency rel.acum.

1	2567	38.05%	38.05%	*****
6	4180	61.95%	100.00%	*****

We can assume that in the education sector the female exceeds the male gender by 29.90%. The mode is female gender. Regarding this fact we could conclude that there may be an increasing feminization over other previous time in which most of the teachers were males.

2.1.2 *Age:*

We have chosen the age variable since it is interesting to note what the average age of employees in the sector of education is. Also the choice of the age variable may help us to see if an aging in the workers of educational sector exists.

In order to make a better interpretation of the age variable, we divide it into 6 subcategories that are the following ones: "1" which represents people who are less than 19 years old. "2" which represents people who are between 20 and 29 years old. "3" stands for people who are 30 to 39 years old. "4" corresponds to 40-49 years. "5" to people aged between 50 and 59 years old. Finally "6" corresponds to people who are more than 59 years old.

Frequency distribution of ANOS2, observations 1-6747

frequency rel. acum.

1	3	0.04%	0.04%	
2	926	13.72%	13.77%	****

3	2223	32.95%	46.72%	*****
4	2078	30.80%	77.52%	*****
5	1174	17.40%	94.92%	*****
6	343	5.08%	100.00%	*

The mode corresponds to the people who are aged between 30-39 years old representing the percentage 32.95% of the total.

We can see that there is a big difference between young adults, those with less than 19 years and those with an age between 20-29 years old, which in total represent a 13.77% in comparison with people who are more than 30 years old. This fact could advise us that there is a possible aging in the education sector. 86, 23% of workers are older than 30 years, it could happen that most of them are individuals with further experience in the sector. People below this age are recent graduates with minimal experience which could explain the low percentage of this kind of subcategory.

All this evidence a possible need for a political and educational measure in order to renew the educational staff in a medium or short term.

2.1.3 Control:

Then we will analyze the differences between what is public and what is private, which means, public or private sector.

The public sector is represented by the number 1, while the private sector is represented by the sector 2.

Frequency distribution of CONTROL, observations 1-6747

frecuencia rel. acum.

1	4199	62.24%	62.24%	*****
2	2548	37.76%	100.00%	*****

We find that majority control is the Public. Today the private sector represents 37.76%, while the public sector is represented with 62.24%.

Currently there are several open questions about the issue of control in the education sector raising several debates about it. The assumption that there is an accelerated

privatization in the Spanish education sector is one of the issues as we have previously stated that causes debate among the Spanish population.

If we assume that there is an increasing privatization of the education sector, the privatization process may have been influenced by the rise of large cities in which there are more polarized societies and middle classes tend to escape from the institutions that enroll the most vulnerable sectors. However, in recent years, this could have been accentuated by the constant arrival of immigrants.

Another possible contributing factor to note is the cultural and economic gap between the small towns and rural Spain with respect to large urban cities where it might happen that the weight of public education is higher. However, on the other hand people tend to think that the increasing privatization is greater when higher income combined with increased urbanization.

Also it is important to note that Spain has been a country where the presence of the Catholic Church has had a great influence on the educational sector. Most private schools belong to the Church while others, also private; belong to associations of parents or secular schools that prefer to get away from centers of education offered by the public school system.

Today, the education sector seems to be privatized while the public sector increasingly shows its weaknesses produced by cuts that often are suffering lately and its special attention to meet the needs of the most vulnerable classes.

2.1.4 *Type of employment contract:*

The type of contract nowadays in a crisis situation in which we live is a very important data to analyze. “1” means an indefinite contract while “2” corresponds to fixed-term contract.

We can suppose that there is an increase in fixed-term contracts whose reasons may come explained by the period of crisis in which we are living. These days where companies and institutions looking to reduce costs, a permanent contract involves an expenditure of larger scale than a fixed-term contract. All of this is due to the compensation to the employee subsequently retired that the company has to pay to him or her. This compensation is a huge cost that the company has to assume. So having written this, it is better for the

company to hire an employee with a fixed-term contract as it is cheaper than the compensation paid to the employee retired.

In addition it should be mentioned the employment policy carried out by the Spanish government which consists in creating fixed-term contracts in order to follow the guidelines set by the European Union based on promoting the creation of fixed-term contracts.

Frequency distribution for TIPOCON, observations 1-6747

Frequency rel.acum.

1	3746	55.52%	55.52%	*****
2	3001	44.48%	100.00%	*****

We can see that the two percentages are similar but indefinite contract exceed 11.04% in comparison with fixed-term contracts. The mode would be 1 which means contract of indefinite duration.

2.1.5 Nationality:

Currently, there is an increase in the establishment of bilingual schools built in Spain which has led to an increase in demand for foreign teachers including native English-speaking, French-speaking and German-speaking.

The selected variable known as TIPOPAIS means the nationality of the individuals salaried in the field of education. It is designated by the number “1” the people who own Spanish nationality whereas the number “2” represents nationality worldwide.

Frequency distribution for TIPOPAIS, observations 1-6747

Frequency rel.acum.

1	6479	96.03%	96.03%	*****
2	268	3.97%	100.00%	*

It is observed that despite the growing demand for English native speakers or other nationality, Spanish nationality (number 1) is the mode with a relative frequency of 96.03% whereas other nationalities (number 2) are represented by a 3.97%.

2.1.6 *Type of study*

Another variable to examine is the variable that represents the kind of education that individuals employed in the education sector own. This variable ranges from “1” represented an individual with no education, passing by the “2” which represents primary education, “3” corresponds to secondary school I, “4” is related with Secondary Education II, “5” represents Vocational medium grade, “6” corresponds to Vocational Degree, “7” is equivalent to College Graduated or equivalent, and finally “8” means Graduates, senior engineers and doctors.

Frequency distributions for ESTU, observations 1-6747:

frecuencia	rel.	acum.	
1	20	0.30%	0.30%
2	167	2.48%	2.77%
3	375	5.56%	8.33% **
4	613	9.09%	17.42% ***
5	179	2.65%	20.07%
6	319	4.73%	24.80% *
7	1607	23.82%	48.61% *****
8	3467	51.39%	100.00% *****

Graduates, senior engineers and doctors account for 51.39% of employees followed by those who are college graduates. Everything is logical because education is currently considered a profession of great social relevance and it makes sense that it is demanded for their professions high levels of both education requirement and access to such education.

2.2 **Descriptive analysis of quantitative variables.**

2.2.1 *Base salary*

The base salary is the salary, paid to a person, which is fixed per unit of time worked or work unit that takes into account both pays and the service provided by others.

The base salary is listed on the payroll with salary bonuses, overtime ... etc.

Main statistical, using observations 1-6747 for variable SALBASE (6747 valid observations):

Media	Mediana	Mínimo	Máximo
-------	---------	--------	--------

1109.97	983.760	0.00000	8532.66
Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
647.430	0.583284	1.27013	5.42427
Porc. 5%	Porc. 95%	Rango IQ	Observaciones ausentes
218.114	2289.03	826.800	0

The mean of the base salary is 1109.97 euros while the median that means, the observation which occupies the central place of the set of observations is 983.760 euros per month

The mode has a lower value in comparison to the median and the mean. Its value is 914.21 euros.

The mean, median and mode are different, being the median higher than the mean and this in turn more than mode. All this means that the distribution is positively skewed or to the right producing that data distribution is asymmetric. This occurs because there are a small proportion of high values. As a result the mean of these distributions is higher than the median. The mean is inflated by people who perceive a high salary which gives a more optimistic view of economic welfare. In this case in which the distribution is positively skewed the median is preferred to the mean in order to describe the central tendency.

All this said before it is confirmed seeing the figure of the box plot in the page X. The median is located near the lower extreme, with a tail extended to the right.

Looking at the frequency distribution, salaries between 761.84-1066.6 euros/month wages are more repetitive with a percentage of 20.25%. In second place, from 1066.6 to 1371.3 euros per month is represented with a percentage of 18.67%. Furthermore, the lower wages euros 2285.5 representing a total of 94.95% of the total. From this wage to wages which are higher, the percentage of the frequency is very small in comparison with other intervals. From this wage all other intervals represent a very small percentage. Only 5.05% represents the wages which are over 2885.5 euros per month.

Great variability exists among base salaries; wages are not homogeneous and vary widely with respect to the mean. This is demonstrated by the high value which is indicated by the standard deviation. This is reaffirmed by the value that the coefficient of variation indicates (approximately 0.58).

Higher kurtosis implies a greater concentration of data near the mean of the distribution simultaneously coexisting with a relatively high data frequency far away from the same. This

explains a form of frequency's distribution with tails very high and with a highly pointed center. In turn, the kurtosis is greater than zero. This causes that the distribution is leptokurtic.

2.2.2 *Years old worked*

Years old worked refers to the period of time in the employment or service. The law takes steps to ensure job stability and establishes benefits that reward the worker's seniority in service and protect them in case of unemployment.

Many people believe that social benefits and seniority benefits are similar concepts. However, they are two quite different items. Social benefits are benefits which include holidays, salary retained and so on. On the other hand, the benefits of seniority are type of savings which the worker is making even when the employment relationship ends. This feature has a social purpose.

Then the report is focused on the descriptive analysis of the year-old worked.

Main statistical, using observations 1-6747 for variable ANOANTI (6747 valid observations)

Media	Mediana	Mínimo	Máximo
8.77012	5.00000	0.00000	48.0000
Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
9.15637	1.04404	1.25619	0.978527
Porc. 5%	Porc. 95%	Rango IQ	Observaciones ausentes
0.00000	28.0000	11.0000	0

The year-old worked's means is 8.7701 while the median is 5 years old.

The mode is 1.71 years old.

The mean is greater than the median and mode. In this case we find that the distribution of the data is biased positively or also called, skewed to the right leaving a tail extending to the right side, in the direction of the highest positive values.

This fact has an explanation because there are a small proportion of high values. As a result the mean of these distributions is higher than the median. The mean is inflated by people who have a very high antiquity which gives a vision that does not accurately reflect reality.

In this instance in which there is a positive skew and specifically the average median is preferable as a measure to describe the central tendency.

All this said before we can display in the box plot figure added in Annexes, N° 2. The median is located near the lower end, reflecting a positive skewed distribution with a tail elongated to the right. We can observe also how outliers are extending along the highest years-old worked.

Noting the frequency distribution we find that the range of years old that have more relative frequency are the years in the range 0.85714 - 2.5714 with a percentage of 19.76%. Followed it is found the range of <0.85714 years old represented by 13.78%.

We may conclude that despite the age of the people is high and that most contracts are undefined, seniority is not as high as them.

There may be several explanations for this. One reason could be that people who incorporated as fixed teachers are people whose age is bigger than 30 years old but are incorporated since just a few years ago or recently, which may explain what it has already explained in the qualitative variable "age" that there is an aging in the sector of the education.

Also it could be another reason that the majority of working people in the education sector are teachers, doctors and staff who have had to make an exam called "oposiciones" in order to get those jobs which takes a lot of time of preparation. Most of people do not pass this exam at first so it takes time also to pass the exam and finally get this kind of job. The second explanation in which we base on, it corresponds to the increase of replacement jobs in this sector.

As we have previously explained, fixed-term contracts have increased occupying 44.48% relative frequency in the data. All this will be analyzed and subsequently explained in the next section.

The coefficient of variation is greater than one reflecting a hyper exponential distribution which is considered as high variance. This aspect is also observed due to the standard deviation (9.15637) that is higher than the mean in general.

Through kurtosis we can identify that there is a high degree of concentration around the core values of the variable which is referred to as leptokurtic distribution.

2.2.3 *Salary complements*

Generally, the salary paid to workers at the end of the month consists of two main concepts: basic salary and salary complements.

The base salary is not set by the company, but each professional category has fixed it in a collective agreement. Why, then, each of these people may have a different salary at the end of the month?

The answer lies in salary complements. These complements make the difference in salary between workers belonging to the same professional and working in the same sector.

There are kinds of complements:

1) Personals: Complements that are derived from the conditions of the worker, as their seniority or their special skills (languages, degrees ...). Also it is included what the company wants to introduce on their own to improve their employees' final salary.

2) Jobs: Bonuses that are related to some positions because of their dangerousness, the involvement of working at night, because they carry a special responsibility among many.

3) Quality and / or quantity of work: performance incentives, incentives for achieving its objectives, punctuality... Here it is also included the overtime.

4) Expiring in over a month: This concept refers to the extra payments and participations in the profits of the company.

5) In "especie": everything the worker receives different from money such as various goods, cars, buildings..etc.

Having explained the concept we will now consider the selected variable. The reason why we have selected this variable is due to the high variability may exist between the salaries of two people working in the same industry because they have different salary complements. It is a good measure of wage inequality that exists in this sector.

Focusing on the descriptive analysis:

Main statistical, using observations 1-6747 for variable Comsal (6747 valid observations)

Media	Mediana	Mínimo	Máximo
618.601	380.700	0.00000	9482.86
Desv. Típica.	C.V.	Asimetría	Exc. de curtosis

739.594	1.19559	2.82471	17.3880
Porc. 5%	Porc. 95%	Rango IQ	Observaciones ausentes
0.00000	1805.96	1013.90	0

The mean wage complements is set to 618.601 euros per month, while the median reflects a lower value, 380.700 euros per month.

The modal class is the smallest interval of 169, 34 euros per month.

The mean is greater than the median and mode. This indicates that there is an asymmetry in the distribution of data, especially positive asymmetry or a skewed distribution to the right.

All this in turn is confirmed because of the asymmetry coefficient which is greater than zero (2.824). It indicates that there is a positive skewed distribution so there is greater concentration of values to the right of the mean than to his left.

The explanation of the bias is because there are a low proportion of high values. As a result the mean of these distributions is higher than the median. The mean is inflated by people who have higher salary complements and this fact gives a vision that does not accurately reflect reality by offering a more positive outlook.

Also analyzing the box plot added in the Annexes, the line represents the median, it indicates the possible symmetry. This line approaches the first quartile indicating that the distribution is skewed to the right.

The frequency distribution indicates that wage supplements with the highest percentage (37.39%) are between 0 and 169.34 euros per month. In second place with a share of 18.33% the wages complements 169.34 and 508.01 euros per month. In the third place we find the wage supplements with values vary from 508.01 to 846.68 euros per month which represent a 12.82% of the total. These low values fully represent 68.55% while over 9313 euros salary complements represent only 0.01%. We can see that there is a high concentration of low values with respect to outliers, all high.

The coefficient of variation is positive and greater than one (1,195) for what is called "high variance". A higher Coefficient of variation only reflects a higher heterogeneity of the variable's values.

The coefficient of kurtosis is greater than 1 (17,388) making it a leptokurtic distribution, that means, it has a high degree of concentration around the core values of the variable.

2.3 Relationship of qualitative variables

2.3.1 Sex and type of study:

We will analyze to what degree gender influences the type of studies which have salaried people in the education sector.

The crosstab added in annexes, n° 1, reports that the number of women is higher in all codes of studies and the reason of that is because there are more women than men in this sector. The code of studies which shows most difference is in the primary studies. There are 64% more women than men.

Usually the difference is high except in the case of graduate degree, senior engineer or doctor where the difference is only 8.8%.

Focusing on women, 45.15% of them belong to the code of degree "graduates, senior engineers or doctors." Then 28.3% of women have a university degree or something similar. And thirdly 8.8% of women have secondary education II title.

Compared with women, men show the same patterns. First, 61.6% of men hold a Bachelor's degree, senior engineer or doctor. Secondly 16.5% of men have a university degree or something like that. And in third place, 9.5% of men have completed their secondary education II.

In relative terms, we see that the distribution of the data focuses primarily on code of studies "8" in the male data (graduate, senior engineer or doctor) leaving small values in other degrees, while in the female data are more distributed among different qualifications.

2.3.2 Age and type of contract:

We will analyze how these two variables are related and see if the hypothesis that people with lower ages have to be familiar with the fixed-term contract. And of course if in opposing, the people with ages are those that are associated with the types of contract of indefinite duration.

It is noted that from age 40 to so on, the percentage of people who have a type of permanent contract is higher than those with a fixed-term contract. For example people

who are 50 to 59 years and own a permanent contract represent 71.9% while those with a fixed-term contract represent 28.1%.

In contrast to this, people under the age of 40 years old represents a higher concentration of fixed-term contracts. If we take a view at the range from 20 years to 29 years, 31% of them hold a permanent contract while 69% hold temporary contracts.

All this leads us to conclude that today as in many areas, the practice of establishing fixed-term contracts, or so-called temporary, among young people is an increasing performance and is accentuated as younger is the population to which they are directed.

In contrast, the older person is salaried, the probability of that person for a permanent contract increases.

The consequence of this could be the continued aging in the population that is dedicated to the education sector as they are not hiring young people indefinitely and still maintain permanent contracts of people whose age is over 40 years old not giving any opportunity for a renewal of teaching.

2.3.3 Nationality and type of control:

Analyzing the crosstab number 3 added in annexes, we conclude that in both aspects of control, both public and private, the Spanish nationality is majority over other nationalities.

In the public sector the Spanish nationality represents 97.4% versus 2.6% representing other nationalities. In the private sector follows the same trend, 93.7% are Spanish nationality while 6.3% is different nationality.

However, we see in the total population that inside of another nationality, working in the field of education, 59.75% represents foreign people who work in the private sector. Moreover, the population who have Spanish nationality, most people of them work in the public sector. It is reflected with a percentage of 63.1% in comparison to the percentage of the private sector (36.9%).

This trend of hiring people of other nationalities in the private sector can be explained by different factors. The first is due to the progressive increase by private schools as bilingual institutions formalized. The second is referred to the need to find native people to school classes or schools of English, French or another language.

In turn we can conclude that the presence of people with Spanish nationality in the public sector can be linked to the different calls for “oposiciones” to the government or the state

convened over the years to get a job as a teacher, administrator, which means to be an officer in general.

2.4 Relationship between quantitative variables:

2.4.1 *Base salary and salary's complements.*

There is a relationship between wages and wage supplements. In this instance it is a negative linear relationship as $r = -0.0143$.

It is a value that is close to 0 which could mean the relationship is weak even it could be a non-linear relationship between base salary and salary complements. But that does not necessarily mean that there is no connection between them.

Coefficient of correlation, using the observations 1 - 6747

Criticalvalue at 5% (a dos colas) = 0.0239 para $n = 6747$

COMSAL	SALBASE	
1.0000	-0.0143	COMSAL
	1.0000	SALBASE

Noting the XY scatter plot added in the annexes, graphic No. X, we see a decreasing slope indicating that there is a negative linear relationship.

2.4.2 *Base salary and years worked.*

There is a relationship between the base salary and years of service of workers in the education sector.

In this case it is a positive linear relationship as $r = 0.1286$.

Coefficient of correlation, using the observations 1 - 6747

Criticalvalue at 5% (a dos colas) = 0.0239 para $n = 6747$

ANOANTI	SALBASE	
1.0000	0.1286	ANOANTI
	1.0000	SALBASE

2.4.3 *Years worked and Salary complements:*

There is a relationship between seniority and wage complements.

In this instance is a positive linear relationship due to the value which the correlation coefficient shows, $r = 0.3241$.

Coefficient of correlation, using the observations 1 - 6747

Criticalvalue at 5% (a dos colas) = 0.0239 para $n = 6747$

ANOANTI	COMSAL	
1.0000	0.3241	ANOANTI
	1.0000	COMSAL

Noting the XY scatter plot added in the annexes, graphic No 3, we can see that there is an increasing positive slope indicating that it is clear there is a positive linear relationship.

2.5 Inferential analysis:

2.5.1 *1st contrast of hypothesis*

H0: $E(SH) = E(SM)$

H1: $E(SH) \neq E(SM)$

Where $E(SH)$ is the average population of the average wage of men and $E(SM)$ is the population mean of the average wage of women.

Hipótesis nula: Diferencia de medias = 0

Muestra 1:

$n = 2567$, media = 1105.05, d.t. = 693.939

desviación típica de la media = 13.6965

Intervalo de confianza 95% para la media: 1078.19 a 1131.91

Muestra 2:

$n = 4180$, media = 1095.4, d.t. = 616.767

desviación típica de la media = 9.53966

Intervalo de confianza 95% para la media: 1076.7 a 1114.1

Estadístico de contraste: $t(6745) = (1105.05 - 1095.4)/16.2293 = 0.594603$

valor p a dos colas = 0.5521

(auna cola = 0.2761)

The null hypothesis is not rejected since p-value equals 0.5521 and is greater than the significance level 0.05, so we can assume that there are no significant differences between the wages paid to men in regard to wages received by women ,there seems not to be a wage gap in general terms (which may be in another subgroup).

Noting the boxplot of the base salary regarding the sex, in the page 10 of the Annexes the salary of men has more atypical case then the one paid to women. That could explain the maximum salary of men which stands in 8532.66 euros per month in comparison to the maximum of women which stands in 4766, 84 euros per month.

We can also see that both box plots have similar dispersion of data.

However the median line in the men's diagram is more focused on the middle of the box that the diagram of the female wage. This indicates that the median female wages are lower than men's wages.Both distributions are positively skewed.

There is a lower coefficient of variation in the wages of men with respect to the coefficient of variation of female wages.However, it remains high and with the corresponding coefficient of variation belonged to the female wage show that both sexes show heterogeneous data.

2.5.2 *2nd Contrast of Hypothesis*

Ho: $E(SCPU) = E(SCPRI)$

H1: $E(SCPU) \neq E(SCPRI)$

Where $E(SCPU)$ is the average population of the average wage of workers in the public sector, and $E(SCPRI)$ is the population mean of the average wage of workers in the private sector.

Hipótesis nula: Diferencia de medias = 0

Muestra 1:

$n = 2548$, media = 1176.73, d.t. = 744.077

desviación típica de la media = 14.7407

Intervalo de confianza 95% para la media: 1147.83 a 1205.63

Muestra 2:

$n = 4199$, media = 1069.46, d.t. = 577.342

desviación típica de la media = 8.90964

Intervalo de confianza 95% para la media: 1051.99 a 1086.93

Estadístico de contraste: $t(6745) = (1176.73 - 1069.46)/16.207 = 6.61876$

valor p a dos colas = 3.901e-011

(auna cola = 1.95e-011)

The p-value is 3.901e-011 which is less than the significance level (0.05) so we reject the null hypothesis. The means of public as well as private wages are similar in meaningful terms. So we can assume that there is a wage gap between the public or private control overall.

This is reinforced by looking at the diagram added in annexes on page 12 made by separation factor “control”. Public control is less spread of data reaching its maximum at 4766.84 euros per month, which coincides with the maximum of female. Whereas the box plot of the private sector is more dispersed and more variable than the public. Curiously the maximum salary in the private sector (8532.66 euros per month) coincides with the maximum male wage.

2.5.3 *3rd Contrast of hypothesis*

$H_0: E(\text{SCD}) = E(\text{SCI})$

$H_1: E(\text{SCD}) \neq E(\text{SCI})$

Where $E(\text{SCD})$ is the average population of the average wage of workers with a fixed-term contract and $E(\text{SCI})$ is the population mean of the average wage of workers with a permanent contract.

Hipótesis nula: Diferencia de medias = 0

Muestra 1:

$n = 3001$, media = 425.99, d.t. = 541.585

desviación típica de la media = 9.8863

Intervalo de confianza 95% para la media: 406.605 a 445.375

Muestra 2:

$n = 3746$, media = 772.906, d.t. = 834.757

desviación típica de la media = 13.6388

Intervalo de confianza 95% para la media: 746.166 a 799.646

Estadístico de contraste: $t(6745) = (425.99 - 772.906)/17.621 = -19.6876$

valor p a dos colas = 6.045e-084

(auna cola = 3.023e-084)

The null hypothesis that the average wages of temporary and permanent contracts are similar in significant terms is rejected. So we can assume that there is a wage gap between certain and permanent contracts in general terms as the p-value is 6.045e-084 less than the significance level.

3.MODEL'S ESTIMATION.

3.1 Dummies

In this third part is carried out the estimation of econometric model. For this we select the qualitative and quantitative variables that we consider most representative.

First, we define a new variable called "L_SALBASE" which represents the logarithm of base salary. The logarithm of this variable is used because of the existence of extreme values, its asymmetry and high dispersion. Therefore the use of logarithms is a tool to soften the distorting effect that means reduce dispersion and simplifying the interpretation of the model.

It is observed that the sample population is high so it is applicable the Central Limit Theorem, so it is reliable to make reliable inference.

To simplify the econometric model is desirable to modify any variable as is the case of the studies level. There are 8 subcategories and therefore we group and divided into two subcategories: 1) People who have a lower level of secondary education II; 2) people who have a higher level of studies than secondary education II. The same it happens with the variable age which is divided between those ones who are less tan 39 years old and those one who are 40 years old or older than this age.

These are the following dummy variables used in the model:

DCONTROL_1=1 If the employee works in a company public or private.

DTIPOCON_1=1 If the worker has a permanent contract or temporary contract.

DSEXO_1=1 if the employee is a male or female.

DTIPOP AIS_1=1 if the employee is Spanish or has another foreign nationality.

DANOS_1=1 If the worker is younger than 39 or older than 40 years.

DESTU2_1=1 If the worker has a level of study less than or equal to Secondary Education II or if the worker has a higher level than Secondary Education Studies II studies.

The model presents following equation:

LSALBASE=

$$\beta_1 + \beta_2 DCONTROL_1 + \beta_3 DTIPOCON_1 + \beta_4 DSEXO_1 + \beta_5 DTIPOP AIS_1 + \beta_6 DANOS_1 + \beta_7 DESTU2_1 + \beta_8 ANOANTI + \beta_9 COMSAL + \mu$$

Using qualitative and quantitative variables set out in paragraph 2, it has estimated the following econometric model to know which of these affect the base salary.

Modelo 2: MCO, using the observation 1-6747 (n = 6746)

It has been removed absent or incomplete observations: 1

Dependen variable: l_SALBASE

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
const	6.44094	0.0452108	142.4645	<0.00001	***
DCONTROL_1	0.106277	0.0186475	5.6993	<0.00001	***
DTIPOCON_1	0.367421	0.0194622	18.8788	<0.00001	***
DSEXO_1	0.00908214	0.0167351	0.5427	0.58735	
DTIPOP AIS_1	-0.0621275	0.0416782	-1.4906	0.13610	
DANOS_1	0.223165	0.0193957	11.5059	<0.00001	***
DESTU2_1	-0.27136	0.0219583	-12.3580	<0.00001	***

ANOANTI	0.0100035	0.00114636	8.7263	<0.00001	***
COMSAL	2.18432e-05	1.22507e-05	1.7830	0.07463	*
Media de la vble. dep.	6.813108	D.T. de la vble. dep.	0.698056		
Suma de cuad. residuos	2940.039	D.T. de la regresión	0.660607		
R-cuadrado	0.105479	R-cuadrado corregido	0.104417		
F(8, 6737)	99.30059	Valor p (de F)	6.5e-157		
Log-verosimilitud	-6770.792	Criterio de Akaike	13559.58		
Criterio de Schwarz	13620.93	Crit. de Hannan-Quinn	13580.76		

3.2 Heteroscedasticity

We can sense that there is heteroscedasticity because the data are not homogeneous.

To check if there is really heteroscedasticity, it is used the White test.

H₀: Homoscedasticity

H₁: Heteroscedasticity.

Estadístico de contraste: $TR^2 = 1019.204678$, con valor $p = P(\text{Chi-cuadrado}(38) > 1019.204678) = 0.000000$

It is observed that the p value is less than the 0.05 level of significance so the null hypothesis is rejected. It states that there is heteroscedasticity. In order to be able to make inference, the model by Ordinary Least Squares with robust typical deviations is estimated. Modelo 2: MCO, using the observations 1-6747 (n = 6746)

It has been removed absent or incomplete observations 1

Dependent variable: L_SALBASE

Desviaciones típicas robustas ante heteroscedasticidad, variante HC1

<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>
-------------------	---------------------	----------------------	----------------

const	6.44094	0.0519743	123.9254	<0.00001	***
DCONTROL_1	0.106277	0.0193469	5.4932	<0.00001	***
DTIPOCON_1	0.367421	0.0203643	18.0424	<0.00001	***
DSEXO_1	0.00908214	0.0171527	0.5295	0.59649	
DTIPOPAIS_1	-0.0621275	0.0472942	-1.3136	0.18901	
DESTU2_1	-0.27136	0.0203997	-13.3022	<0.00001	***
DANOS_1	0.223165	0.0209497	10.6524	<0.00001	***
ANOANTI	0.0100035	0.00123884	8.0749	<0.00001	***
COMSAL	2.18432e-05	1.00404e-05	2.1755	0.02962	**

Media de la vble. dep.	6.813108	D.T. de la vble. dep.	0.698056
Suma de cuad. residuos	2940.039	D.T. de la regresión	0.660607
R-cuadrado	0.105479	R-cuadrado corregido	0.104417
F(8, 6737)	105.2613	Valor p (de F)	4.4e-166
Log-verosimilitud	-6770.792	Criterio de Akaike	13559.58
Criterio de Schwarz	13620.93	Crit. de Hannan-Quinn	13580.76

3.3 ANOVA Contrast:

Ho: Econometricmodel no significant

H1: Econometricmodelsignificant

F(8, 6737)	105.2613	Valor p (de F)	4.4e-166
------------	----------	----------------	----------

P value is lower than 0.05 the null hypothesis is rejected so the econometric model is globally significant.

3.4 Interpretation of the model econometrics variables.

Control: Workers who work in the public sector corresponding to the education sector receive a wage base, specifically a 10.62% higher than the corresponding bases wages to the workers in private control centers, provided that the other variables remain constant.

Type of contract: Those who have a permanent contract receive a salary base greater by 36.4% to those working under the conditions of a fixed-term contract. All of this is allowed while maintaining the other variables constant.

Sex: Men charged 0.91% more than women. It is observed that there is no wage gap between the sexes. And if there is, this is minimal.

Nationality: According to the value displayed on the model, it is concluded that workers from other countries receive a salary base 6% higher than the national workers.

Type of study: As previously announced at the beginning of this reporting, the qualitative variable type of study has been divided and regrouped, eventually getting two subcategories: 1) People who have a lower level of secondary education II; 2) people who have a higher level of studies than secondary education II.

People who possess a level corresponding to no education, primary education and secondary education have a basic salary 27% lower than those who have intermediate degree level or who are graduates and doctors. All of this is possible maintaining the rest of the variables constant.

Age: The variable years has several categories which have been merged finally obtaining two main categories: 1) People aged between 0-40 years; 2) Individuals aged between 40 and 60 years. It is observed that people under 40 years receive a salary base greater approximately 22.32% than those over 40. This can be explained by people whose age is between 30 and 40 years as it is the age that most of people work in the education sector compared to other ages.

Years worked: For each year of age that education workers have their base salary increased by 1% keeping all other variables constant.

Salary complements: For every euro salary per month that has a worker in education complement his base salary decreases by 2.18432e-03% holding all other variables constant.

This coefficient is very small which means that even if significant, the differences between them would be minimal.

3.5 Individual significance:

Next we will examine if the dummies variables are significant or not to explain the model. We only have selected those which we consider there will be something interesting to mention.

1st Hypothesis:

Ho: β_4 is not significant

H1: β_4 is significant.

DSEXO_1	0.00908214	0.0171527	0.5295	0.59649
---------	------------	-----------	--------	---------

As the p-value is higher than 0.05 (p-value: 0.59649) the null hypothesis is not rejected. At this level of significance we can assume that this dummy called Sex results not significant in order to explain the base pay of the employees in the educative sector. . Se puede asumir que no hay discriminación salarial por sexo.

2nd Hypothesis:

Ho: β_5 is not significant

H1: β_5 is significant.

DTIPOPAS_1	-0.0621275	0.0472942	-1.3136	0.18901
------------	------------	-----------	---------	---------

As the p-value is higher than 0.05, which is the level of significance, the null hypothesis is not rejected meaning that we can assume the dummy called “tipo de país” which represents the nationality of the people is not significant to explain the base pay of the employees in the educative sector.

3 rd hypothesis:

Ho: β_9 is not significant

H1: β_9 is significant.

COMSAL	2.18432e-05	1.00404e-05	2.1755	0.02962	**
--------	-------------	-------------	--------	---------	----

As the p-value is smaller than 0.05 the null hypothesis is reject. At this level of significance there is enough evidence that the dummy called “complementos salariales” is significant or representative in order to explain the base pay of the employees in the educative sector.

The rest of the variables relevant to explain the model the p- value is smaller than 0,05; (p-value: 0,00001) so the null hypothesis will be rejected.

At this level of significant we can assume that the rest of the dummies result relevant in order to explain the base pay of the employees in the educative sector.

3.6 Study of the multicollinearity's problem:

We would like to see if our model presents some kind of multicollinearity's problem that is a situation in which a strong correlation between explanatory variables is presented.

Factores de inflación de varianza (VIF)

Mínimo valor posible = 1.0

Valores mayores que 10.0 pueden indicar un problema de colinealidad

DCONTROL_1	1.263
DTIPOCON_1	1.446
DSEXO_1	1.021
DTIPOPAYS_1	1.024
DESTU2_1	1.072
DANOS_1	1.448
ANOANTI	1.703

COMSAL 1.269

$VIF(j) = 1/(1 - R(j)^2)$, where $R(j)$ is the coefficient of multiple correlation between the variable j and the rest of independent variables.

Properties of the matrix $X'X$:

norma-1 = 6.3414036e+009

Determinant = 9.3314502e+036

Reciprocalconditionnumber= 1.5455243e-008

Once we have found in the corresponding gretl analysis, the results that show us the detection whether or not there is collinearity. In this case, because all of the values which are smaller than 10, there is no problem of collinearity, therefore it does not exist a huge correlation between all the variables. This makes easy the fact of dividing the partial effect in each one of the dummies over the dependent variable.

3.7 Final econometric model

Finally, we could propose an econometric model in which we can see the variables that are really relevant. It is a model less parsimonious (with less variables) and with a similar interpretation

Modelo 1: MCO, usando las observaciones 1-6747 (n = 6746)

Se han quitado las observaciones ausentes o incompletas: 1

Variable dependiente: l_SALBASE

Desviaciones típicas robustas ante heterocedasticidad, variante HC1

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
const	6.3891	0.0278543	229.3755	<0.00001	***
DCONTROL_1	0.114454	0.0183248	6.2459	<0.00001	***
DTIPOCON_1	0.374444	0.0204072	18.3486	<0.00001	***
DESTU2_1	-0.277977	0.0197882	-14.0476	<0.00001	***

DANOS_1	0.221359	0.0209906	10.5456	<0.00001	***
ANOANTI	0.0102491	0.00121826	8.4129	<0.00001	***
Media de la vble. dep.	6.813108	D.T. de la vble. dep.	0.698056		
Suma de cuad. residuos	2942.704	D.T. de la regresión	0.660759		
R-cuadrado	0.104668	R-cuadrado corregido	0.104004		
F(5, 6740)	166.6309	Valor p (de F)	1.4e-167		
Log-verosimilitud	-6773.848	Criterio de Akaike	13559.70		
Criterio de Schwarz	13600.60	Crit. de Hannan-Quinn	13573.82		

In this final model, the coefficients estimators have varied weakly with respect to the initial model. Now the model is more parsimonious the initial one.

4. CONCLUSIONS

Carried out the study, it is uncovered a number of conclusions about the Spanish wage structure and with special attention of the wage structure in the education sector.

As previously noted in the introduction, the determination of salary depends on many variables, both social and economic.

The number of variables affecting the wage structure is high although not all have the same level of effect on her.

To determine which variables are representative and assisting the creation of econometric model, is seen that variables such as gender and nationality does not affect or has not a high effect in determining the salary of a person. However, there are variables such as years worked or type of contracts which affect it considerably.

Another conclusion to note is the continually changing wage structure suffers. Years ago it had no pay structure variables such as paid holidays or overtime because there did not exist or were not taken into account when we established the wage.

Even today there are new variables that were previously unthinkable as the type of nationality since there were not a large number of foreigners working in Spain as is the case

now. Therefore, one could say that in a few years, new social and economic variables will appear affecting the wage and they must be taken into account when we will analyze the wage structure.

Summarizing the main conclusions obtained through the analysis of the data, it could be argued that the Spanish wage structure is differentiated by sector.

There is a high salary range. This applies because those sectors with a high technological level and jobs that require a degree of technological skill, earn higher salary than the average wages.

On the other hand, those workers belonging to weaker segments such as youth, immigrants, and women earn lower wages, they maintain a weak position when receive wages.

In the case of education sector, gender does not affect the wage structure in a representative way so it is not a significant variable and do not influence.

This fact it is explained because the salary of educational sector is established by agreement so it is less likely to appear a discrimination of this kind. However the discrimination made by the level of study is bigger as the level of study increase.

It is noteworthy that the wage gap between the sexes, men and women, could accentuate when the smaller is the knowledge required by the industry in which they work.

In this regard, companies belonging to the sector require a great capacity for knowledge by their workers are the least discriminate or make less differences based on sex.

From my personal perspective, this work has led to me realize how wages are set and their interpretation.

In my opinion, the salary structure should be a topic known for everyone as the Spanish people are unaware that many variables affect this and do not care about it.

It is true that due to the economic situation we are going through and the few opportunities offered to young market difficult the increase of interest to know what really the salary we perceive or will perceive someday is.

The salary generates changes or at least it demands them, and therefore it is of great importance for the development and evolution of a society so we need to take into account that is a determination of measure the quality of life we have.

5. REFERENCES.

Comunidad Escolar (2014). Aumenta la edad media en el profesorado. Recuperado el 5 de Junio de 2014, de [//http://comunidad-escolar.cnice.mes.es](http://comunidad-escolar.cnice.mes.es)

Laboris. Complementos salariales: marcando la diferencia cada fin de mes. Recuperado el 13 de diciembre de 2013, de [//http://www.laboris.net](http://www.laboris.net)

Miguel Á. Ferrer López (2013). Cómo confeccionar nóminas y seguros sociales.(Deusto)

Estadística para todos. Diagrama de caja y bigotes. Recuperado el 10 de octubre de 2008, de [//http://estadisticaparatodos.es](http://estadisticaparatodos.es)

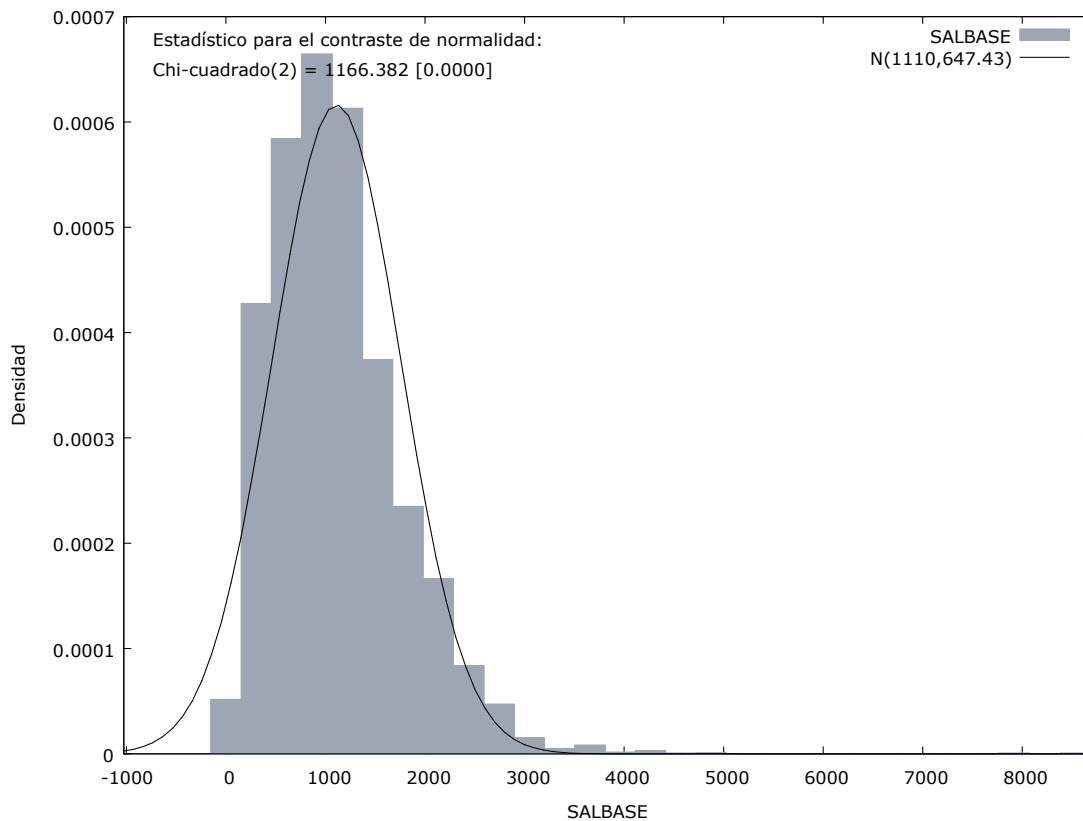
Perfil Sistemas de Información de Universidad de Ciudad Real (2003-2004) Diapositivas 5º Curso de Ingeniería informática.

Carlos J. Perez (2012). Econometría. Guía Rápida de Gretl. (Universidad Carlos III de Madrid).

Paul Newbold, William L. Carlson, Betty Thorne. (2008) Estadística para administración y economía. Madrid. Editorial Pearson Prentice Hall

6 ANNEXES

Graphic n° 1, Base salary:



Frequencydistribution

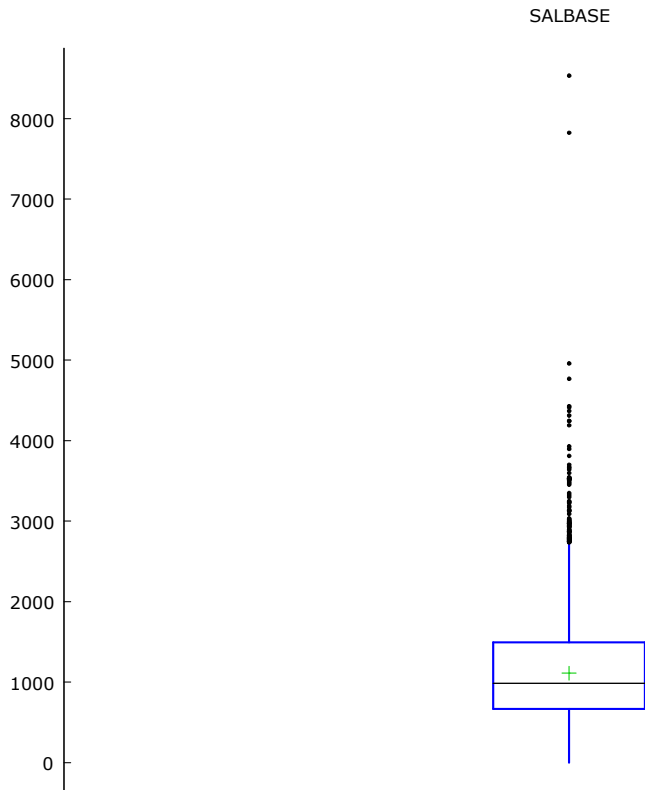
Distribución de frecuencias para SALBASE, observaciones 1-6747
número de cajas = 29, media = 1109.97, desv.típ.=647.43

intervalo	punto medio	frecuencia	relacum.
< 152.37	0.0000	106	1.57%
152.37 - 457.11	304.74	879	13.03%
457.11 - 761.84	609.48	1201	17.80%
761.84 - 1066.6	914.21	1366	20.25%
1066.6 - 1371.3	1219.0	1260	18.67%
1371.3 - 1676.1	1523.7	769	11.40%
1676.1 - 1980.8	1828.4	483	7.16%
1980.8 - 2285.5	2133.2	342	5.07%
2285.5 - 2590.3	2437.9	172	2.55%
2590.3 - 2895.0	2742.6	97	1.44%
2895.0 - 3199.7	3047.4	31	0.46%
3199.7 - 3504.5	3352.1	10	0.15%
3504.5 - 3809.2	3656.9	17	0.25%
3809.2 - 4114.0	3961.6	3	0.04%
4114.0 - 4418.7	4266.3	6	0.09%
4418.7 - 4723.4	4571.1	1	0.01%
4723.4 - 5028.2	4875.8	2	0.03%
5028.2 - 5332.9	5180.5	0	0.00%
5332.9 - 5637.7	5485.3	0	0.00%
5637.7 - 5942.4	5790.0	0	0.00%
5942.4 - 6247.1	6094.8	0	0.00%
6247.1 - 6551.9	6399.5	0	0.00%
6551.9 - 6856.6	6704.2	0	0.00%

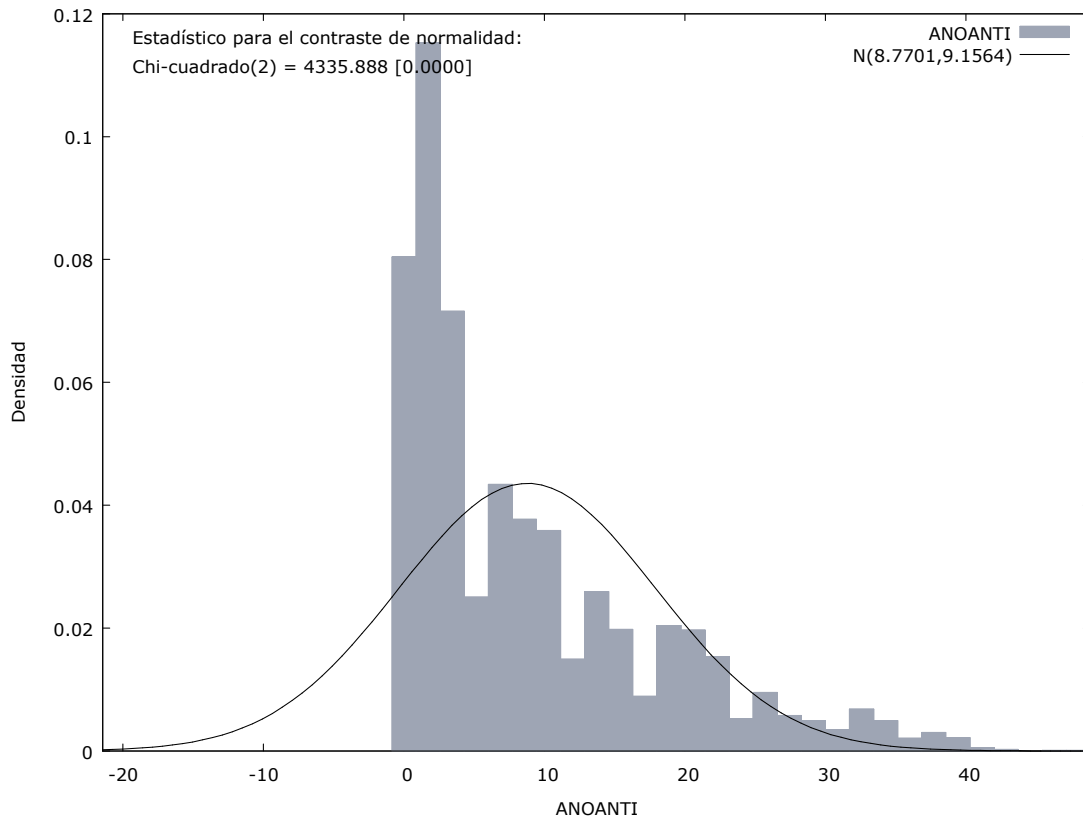
6856.6 - 7161.3	7009.0	0	0.00%	99.97%
7161.3 - 7466.1	7313.7	0	0.00%	99.97%
7466.1 - 7770.8	7618.4	0	0.00%	99.97%
7770.8 - 8075.6	7923.2	1	0.01%	99.99%
8075.6 - 8380.3	8227.9	0	0.00%	99.99%
>= 8380.3	8532.7	1	0.01%	100.00%

Contraste de la hipótesis nula de distribución normal:
 Chi-cuadrado(2) = 1166.382 con valor p 0.00000

Box Plot nº 1 of base salary



Graphic 2 Years-oldworked.



FREQUENCY DISTRIBUTION OF YEARS- OLD WORKED

Distribución de frecuencias para ANOANTI, observaciones 1-6747

número de cajas = 29, media = 8.77012, desv.típ.=9.15637

intervalo punto medio frecuencia relacum.

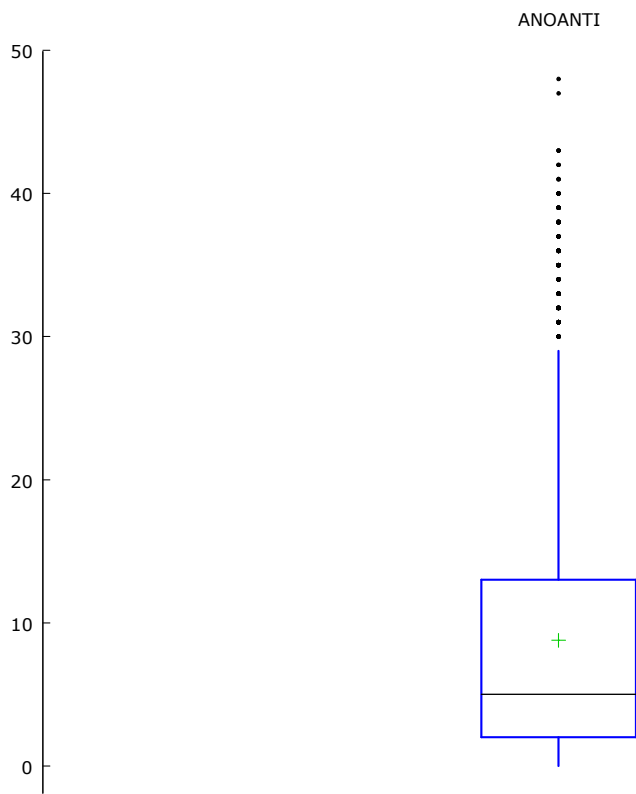
<	0.85714	0.0000	930	13.78%	13.78%	****
	0.85714	- 2.5714	1.7143	1333	19.76%	33.54% *****
	2.5714	- 4.2857	3.4286	828	12.27%	45.81% ****
	4.2857	- 6.0000	5.1429	290	4.30%	50.11% *
	6.0000	- 7.7143	6.8571	502	7.44%	57.55% **
	7.7143	- 9.4286	8.5714	436	6.46%	64.01% **
	9.4286	- 11.143	10.286	415	6.15%	70.16% **
	11.143	- 12.857	12.000	173	2.56%	72.73%
	12.857	- 14.571	13.714	300	4.45%	77.18% *

14.571 -	16.286	15.429	229	3.39%	80.57% *
16.286 -	18.000	17.143	103	1.53%	82.10%
18.000 -	19.714	18.857	236	3.50%	85.59% *
19.714 -	21.429	20.571	228	3.38%	88.97% *
21.429 -	23.143	22.286	178	2.64%	91.61%
23.143 -	24.857	24.000	61	0.90%	92.52%
24.857 -	26.571	25.714	110	1.63%	94.15%
26.571 -	28.286	27.429	67	0.99%	95.14%
28.286 -	30.000	29.143	57	0.84%	95.98%
30.000 -	31.714	30.857	40	0.59%	96.58%
31.714 -	33.429	32.571	79	1.17%	97.75%
33.429 -	35.143	34.286	57	0.84%	98.59%
35.143 -	36.857	36.000	24	0.36%	98.95%
36.857 -	38.571	37.714	35	0.52%	99.47%
38.571 -	40.286	39.429	25	0.37%	99.84%
40.286 -	42.000	41.143	6	0.09%	99.93%
42.000 -	43.714	42.857	3	0.04%	99.97%
43.714 -	45.429	44.571	0	0.00%	99.97%
45.429 -	47.143	46.286	1	0.01%	99.99%
>=	47.143	48.000	1	0.01%	100.00%

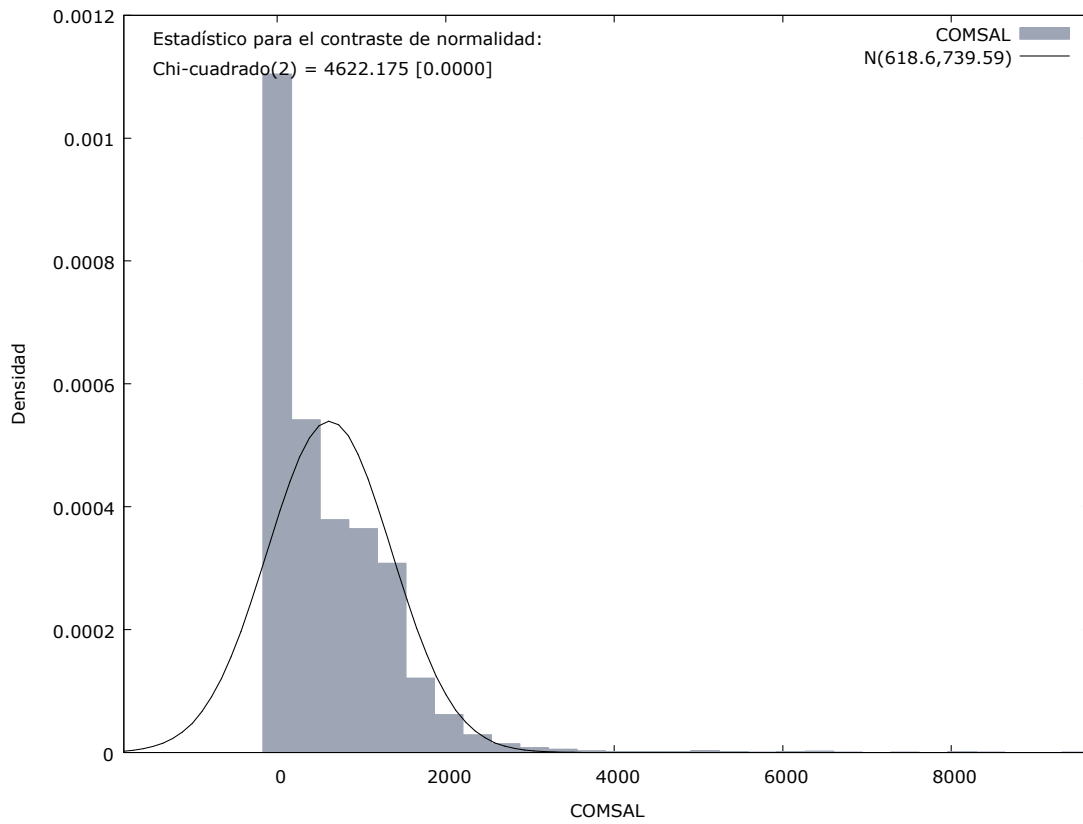
Contraste de la hipótesis nula de distribución normal:

Chi-cuadrado(2) = 4335.888 con valor p 0.00000

Box Plot nº 2 of Years old worked:



Graphic 3rd Salarycomplements.



FREQUENCY DISTRIBUTION:

Distribución de frecuencias para COMSAL, observaciones 1-6747

número de cajas = 29, media = 618.601, desv.típ.=739.594

intervalo punto medio frecuencia relacum.

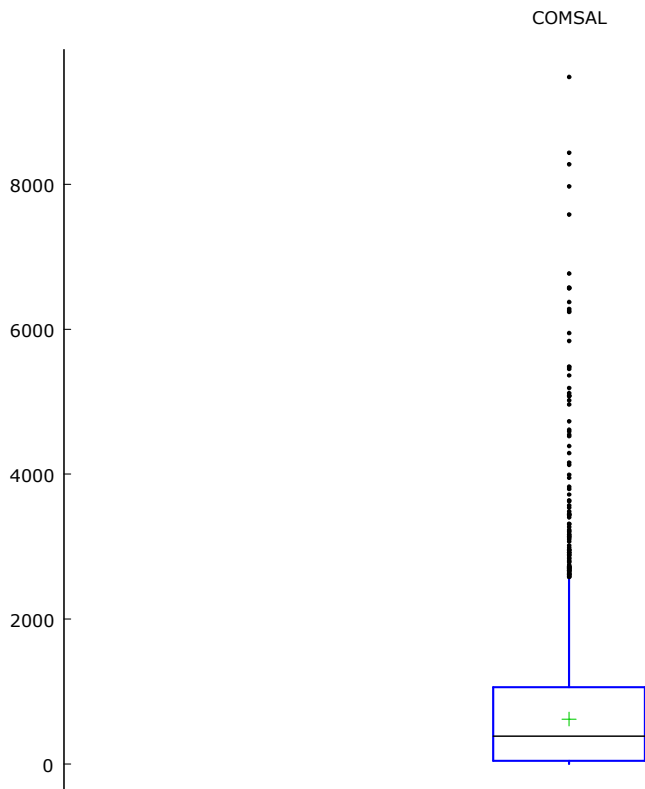
< 169.34	0.0000	2523	37.39%	37.39%	*****
169.34 - 508.01	338.67	1237	18.33%	55.73%	*****
508.01 - 846.68	677.35	865	12.82%	68.55%	****
846.68 - 1185.4	1016.0	832	12.33%	80.88%	****
1185.4 - 1524.0	1354.7	704	10.43%	91.31%	***
1524.0 - 1862.7	1693.4	276	4.09%	95.41%	*
1862.7 - 2201.4	2032.0	140	2.07%	97.48%	
2201.4 - 2540.1	2370.7	65	0.96%	98.44%	

2540.1	-	2878.7	2709.4	32	0.47%	98.92%
2878.7	-	3217.4	3048.1	18	0.27%	99.18%
3217.4	-	3556.1	3386.7	12	0.18%	99.36%
3556.1	-	3894.7	3725.4	6	0.09%	99.45%
3894.7	-	4233.4	4064.1	4	0.06%	99.51%
4233.4	-	4572.1	4402.8	4	0.06%	99.57%
4572.1	-	4910.8	4741.4	3	0.04%	99.61%
4910.8	-	5249.4	5080.1	7	0.10%	99.72%
5249.4	-	5588.1	5418.8	4	0.06%	99.78%
5588.1	-	5926.8	5757.5	1	0.01%	99.79%
5926.8	-	6265.5	6096.1	3	0.04%	99.84%
6265.5	-	6604.1	6434.8	5	0.07%	99.91%
6604.1	-	6942.8	6773.5	1	0.01%	99.93%
6942.8	-	7281.5	7112.1	0	0.00%	99.93%
7281.5	-	7620.2	7450.8	1	0.01%	99.94%
7620.2	-	7958.8	7789.5	0	0.00%	99.94%
7958.8	-	8297.5	8128.2	2	0.03%	99.97%
8297.5	-	8636.2	8466.8	1	0.01%	99.99%
8636.2	-	8974.8	8805.5	0	0.00%	99.99%
8974.8	-	9313.5	9144.2	0	0.00%	99.99%
>=	9313.5	9482.9	1	0.01%	100.00%	

Contraste de la hipótesis nula de distribución normal:

Chi-cuadrado(2) = 4622.175 con valor p 0.00000

Box Plot nº 3 of Salary's complement.



CROSSTABS:

1) SEX AND TIPE OF STUDY

The sex is divided into male and female. The masculine gender is designated by the number 1 and the female gender is designated by the number 6.

The type of study ranges from “1” represented an individual with no education, passing by the “2” which represents primary education, “3” corresponds to secondary school I, “4” is related with Secondary Education II, “5” represents Vocational medium grade, “6” corresponds to Vocational Degree, “7” is equivalent to College Graduated or equivalent, and finally “8” means Graduates, senior engineers and doctors.

Tabulación cruzada de SEXO (filas) contra ESTU (columnas)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	TOT.
[1]	4	30	119	245	48	116	423	1582	2567
[6]	16	137	256	368	131	203	1184	1885	4180
TOTAL	20	167	375	613	179	319	1607	3467	6747

Contraste chi-cuadrado de Pearson = 226.906 (7 gl, valor p = 2.25459e-045)

Tabulación cruzada de SEXO (filas) contra ESTU (columnas)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	TOT.
[1]	0.2%	1.2%	4.6%	9.5%	1.9%	4.5%	16.5%	61.6%	2567
[6]	0.4%	3.3%	6.1%	8.8%	3.1%	4.9%	28.3%	45.1%	4180
TOTAL	0.3%	2.5%	5.6%	9.1%	2.7%	4.7%	23.8%	51.4%	6747

Contraste chi-cuadrado de Pearson = 226.906 (7 gl, valor p = 2.25459e-045)

Tabulación cruzada de SEXO (filas) contra ESTU (columnas)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	TOT.
[1]	20.0%	18.0%	31.7%	40.0%	26.8%	36.4%	26.3%	45.6%	38.0%
[6]	80.0%	82.0%	68.3%	60.0%	73.2%	63.6%	73.7%	54.4%	62.0%
TOTAL	20	167	375	613	179	319	1607	3467	6747

Contraste chi-cuadrado de Pearson = 226.906 (7 gl, valor p = 2.25459e-045)

2) AGE AND TYPE OF CONTRACT.

In order to make a better interpretation of the age variable, we divide it into 6 subcategories that are the following ones: “1” which represents people who are less than 19 years old. “2” which represents people who are between 20 and 29 years old. “3” stands for people who are 30 to 39 years old. “4” corresponds to 40-49 years. “5” to people aged between 50 and 59 years old. Finally “6” corresponds to people who are more than 59 years old.

Type of contract is divided into: “1” means an indefinite contract while “2” corresponds to fixed-term contract.

Tabulación cruzada de ANOS2 (filas) contra TIPOCON (columnas)

	[1]	[2]	TOT.
[1]	1	2	3
[2]	287	639	926
[3]	1065	1158	2223
[4]	1361	717	2078
[5]	844	330	1174

```
[ 6] 188 155 343
```

```
TOTAL 3746 3001 6747
```

Contraste chi-cuadrado de Pearson = 489.534 (5 gl, valor p = 1.44935e-103)

Tabulación cruzada de ANOS2 (filas) contra TIPOCON (columnas)

```
      [ 1][ 2] TOT.  
[ 1] 33.3% 66.7% 3  
[ 2] 31.0% 69.0% 926  
[ 3] 47.9% 52.1% 2223  
[ 4] 65.5% 34.5% 2078  
[ 5] 71.9% 28.1% 1174  
[ 6] 54.8% 45.2% 343
```

```
TOTAL 55.5% 44.5% 6747
```

Contraste chi-cuadrado de Pearson = 489.534 (5 gl, valor p = 1.44935e-103)

3) CONTROL AND NATIONALITY.

The public sector is represented by the number 1, while the private sector is represented by the sector 2. Both variables are located in rows.

The selected variable known as TIPOPAIS means the nationality of the individuals salaried in the field of education. It is designated by the number "1" the people who own Spanish nationality whereas the number "2" represents nationality worldwide.

Tabulación cruzada de CONTROL (filas) contra TIPOPAIS (columnas)

```
      [ 1][ 2] TOT.  
[ 1] 4091 108 4199  
[ 2] 2388 160 2548  
TOTAL 6479 268 6747
```

Contraste chi-cuadrado de Pearson = 57.1414 (1 gl, valor p = 4.05582e-014)

Contraste exacto de Fisher:

Left: P-value = 1

Right: P-value = 9.69113e-014

2-Tail: P-value = 1.51584e-013

Tabulación cruzada de CONTROL (filas) contra TIPOPAIS (columnas)

	[1]	[2]	TOT.
[1]	97.4%	2.6%	4199
[2]	93.7%	6.3%	2548
TOTAL	96.0%	4.0%	6747

Contraste chi-cuadrado de Pearson = 57.1414 (1 gl, valor p = 4.05582e-014)

Contraste exacto de Fisher:

Left: P-value = 1

Right: P-value = 9.69113e-014

2-Tail: P-value = 1.51584e-013

Tabulación cruzada de CONTROL (filas) contra TIPOPAIS (columnas)

	[1]	[2]	TOT.
[1]	63.1%	40.3%	62.2%
[2]	36.9%	59.7%	37.8%
TOTAL	6479	268	6747

Contraste chi-cuadrado de Pearson = 57.1414 (1 gl, valor p = 4.05582e-014)

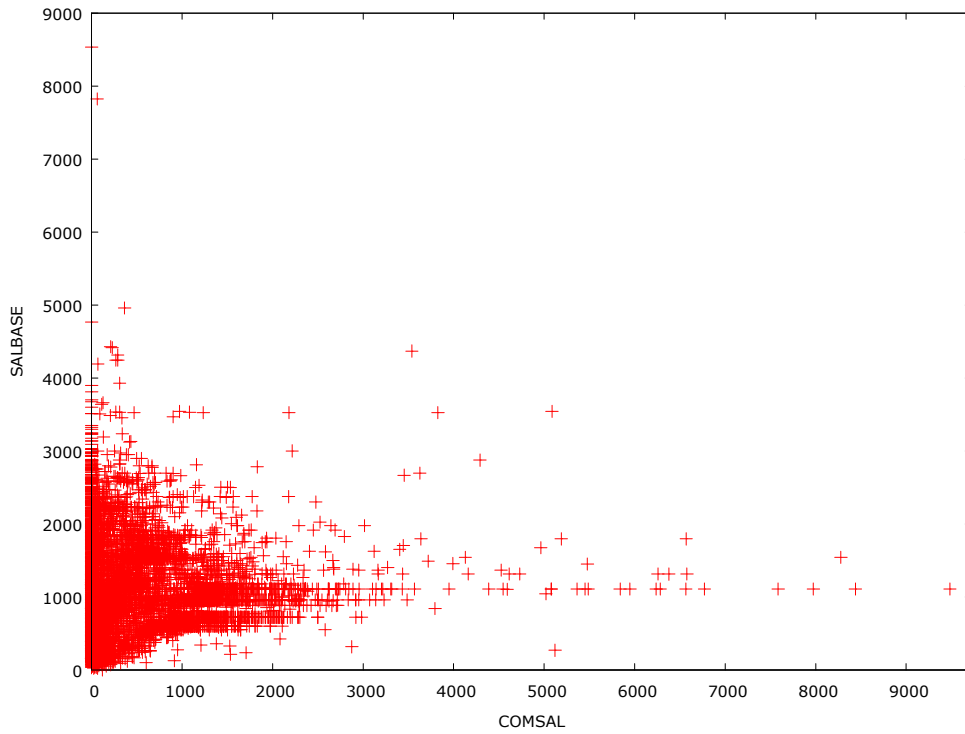
Contraste exacto de Fisher:

Left: P-value = 1

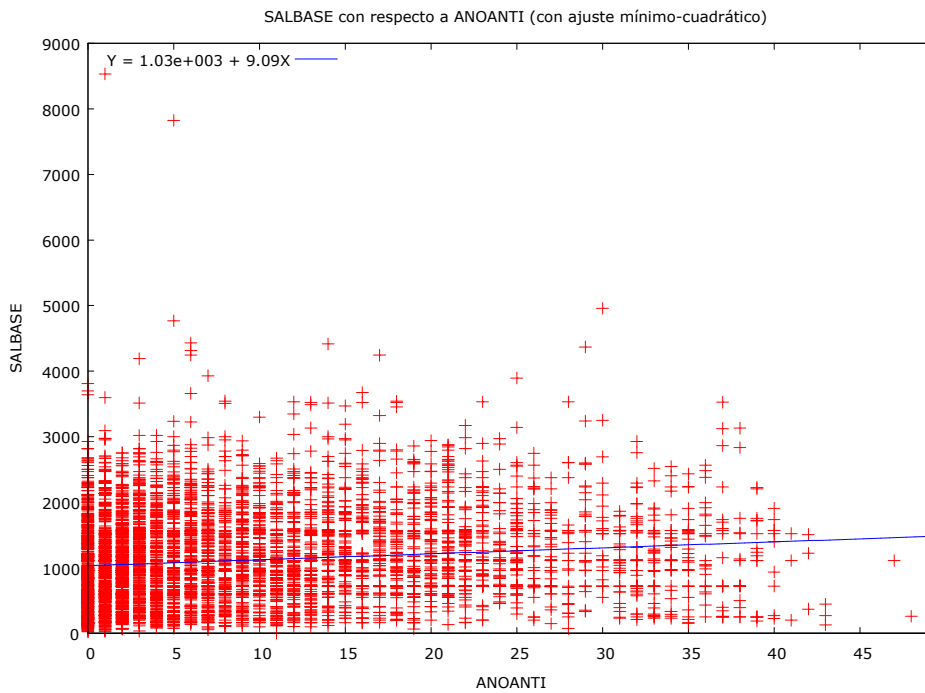
Right: P-value = 9.69113e-014

2-Tail: P-value = 1.51584e-013

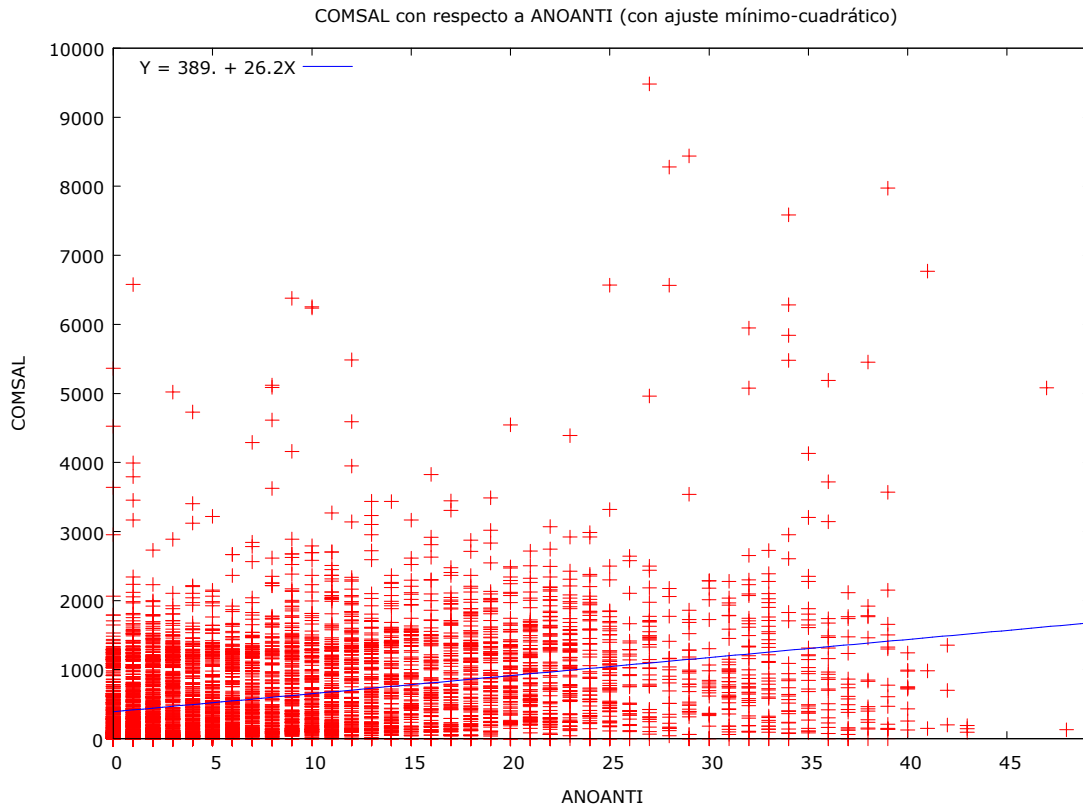
GRAPHIC n° 4 OF CORRELATION BETWEEN BASE SALARY AND SALARY COMPLEMENTS.



GRAPHIC n° 5 OF CORRELATION BETWEEN BASE SALARY AND YEARS WORKED.



GRAPHIC n° 6 OF CORRELATION BETWEEN SALARY COMPLEMENTS AND YEARS WORKED.



CONTRAST OF HYPOTHESIS:

1st contrast of hypothesis.

Contrast between the average wage of men and average wage of women.

H0: $E(SH) = E(SM)$

H1: $E(SH) \neq E(SM)$

Where $E(SH)$ is the average population of the average wage of men and $E(SM)$ is the population mean of the average wage of women.

Hipótesis nula: Diferencia de medias = 0

Muestra 1:

n = 2567, media = 1105.05, d.t. = 693.939

desviación típica de la media = 13.6965

Intervalo de confianza 95% para la media: 1078.19 a 1131.91

Muestra 2:

n = 4180, media = 1095.4, d.t. = 616.767

desviación típica de la media = 9.53966

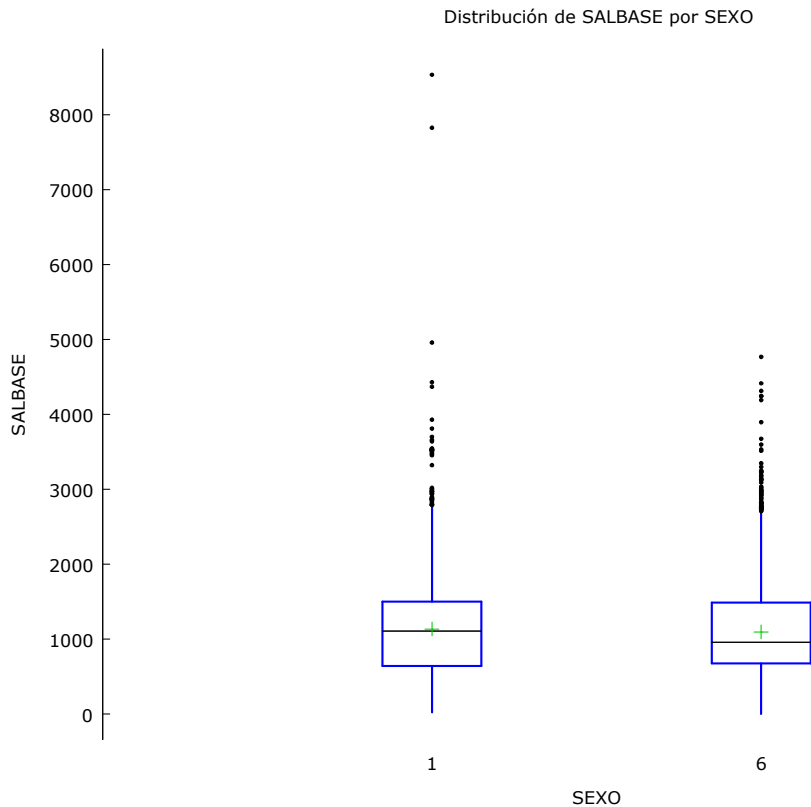
Intervalo de confianza 95% para la media: 1076.7 a 1114.1

Estadístico de contraste: $t(6745) = (1105.05 - 1095.4)/16.2293 = 0.594603$

valor p a dos colas = 0.5521

(a una cola = 0.2761)

BOXPLOT N° 4 SALBASE con factor de separación SEXO.



Noting the boxplot of the base salary regarding the sex, in the page 10 of the Annexes the salary of men has more atypical case than the one paid to women. That could explain the maximum salary of men which stands in 8532.66 euros per month in comparison to the maximum of women which stands in 4766, 84 euros per month.

2nd Contrast of Hypothesis

CONTRAST BETWEEN THE AVERAGE WAGE OF WORKERS IN PUBLIC SECTOR AND AVERAGE WAGE OF WORKERS IN THE PRIVATE SECTOR.

$H_0: E(SCPU) = E(SCPRI)$

$H_1: E(SCPU) \neq E(SCPRI)$

Where E (SCPU) is the average population of the average wage of workers in the public sector, and E (SCPRI) is the population mean of the average wage of workers in the private sector.

Hipótesis nula: Diferencia de medias = 0

Muestra 1:

n = 2548, media = 1176.73, d.t. = 744.077

desviación típica de la media = 14.7407

Intervalo de confianza 95% para la media: 1147.83 a 1205.63

Muestra 2:

n = 4199, media = 1069.46, d.t. = 577.342

desviación típica de la media = 8.90964

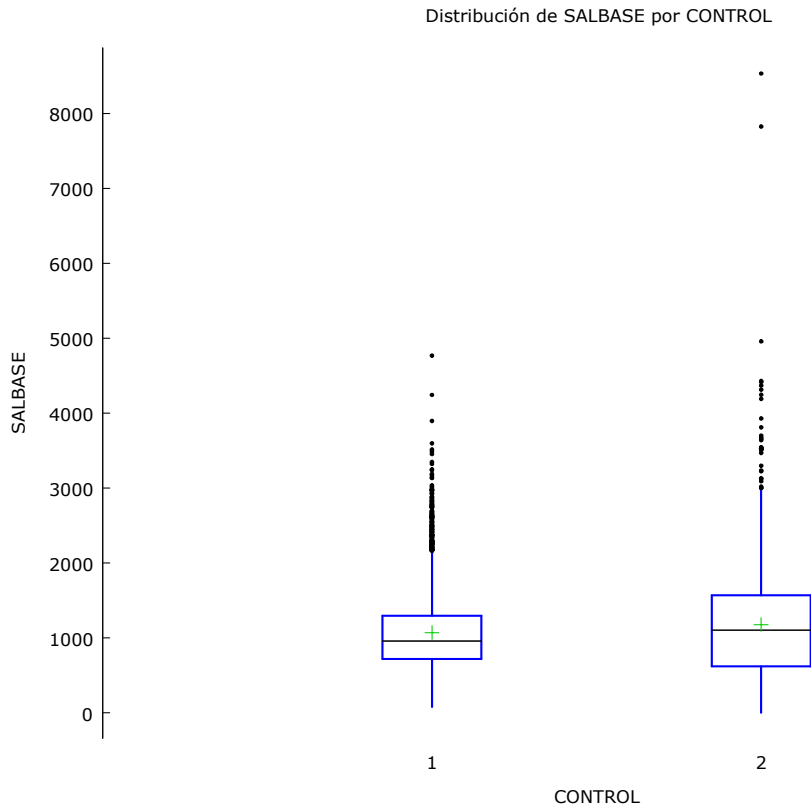
Intervalo de confianza 95% para la media: 1051.99 a 1086.93

Estadístico de contraste: $t(6745) = (1176.73 - 1069.46)/16.207 = 6.61876$

valor p a dos colas = 3.901e-011

(a una cola = 1.95e-011)

BOXPLOT N°5 SALBASE con factor de separación CONTROL.



In this box plot some atypical point are seen. Public control is less spread of data reaching its maximum at 4766.84 euros per month, which coincides with the maximum of female. Whereas the box plot of the private sector is more dispersed and more variable than the public. Curiously the maximum salary in the private sector (8532.66 euros per month) coincides with the maximum male wage.

3rd Contrast of hypothesis

CONTRAST OF THE AVERAGE WAGE OF WORKERS WITH FIXED TERM CONTRACT AND AVERAGE WAGE OF WORKERS WITH A PERMANET CONTRACT.

H0: $E(\text{SCD})=E(\text{SCI})$

H1: $E(\text{SCD})\neq E(\text{SCI})$

Where $E(\text{SCD})$ is the average population of the average wage of workers with a fixed-term contract and $E(\text{SCI})$ is the population mean of the average wage of workers with a permanent contract.

Hipótesis nula: Diferencia de medias = 0

Muestra 1:

n = 3001, media = 425.99, d.t. = 541.585

desviación típica de la media = 9.8863

Intervalo de confianza 95% para la media: 406.605 a 445.375

Muestra 2:

n = 3746, media = 772.906, d.t. = 834.757

desviación típica de la media = 13.6388

Intervalo de confianza 95% para la media: 746.166 a 799.646

Estadístico de contraste: $t(6745) = (425.99 - 772.906)/17.621 = -19.6876$

valor p a dos colas = 6.045e-084

(a una cola = 3.023e-084)

BLOXPLOT N°6 SALBASE Con factor de separación TIPOCON.

