eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

EAP MASTER

FACULTY OF ECONOMICS

Pedro Iraburu Muñoz

LABOR MARKETS:

analysis of Muestra Continua de Vidas Laborales.

13th of April, 2020

Marbella

Analysis of the labor market

Pedro Iraburu Muñoz

13/04/2020

Abstract

This short project will analyze the Muestra Continua de Vidas Laborales (Seguridad Social, 2020) database, by first, computing a panel data, and by studying closely the variables related to the laboral life of the different groups of immigrants within Spain.

Keywords: MCVL, immigrants, Labor Markets.

Introduction

Immigrations has been the focus of the study of many economists. The assimilation of immigrants in the labor market in Spain is a difficult issue to study, as the waves of immigration to Spain are fairly recent. In this assignment, I will analyze the labor life: wages, contract types, laboral situation, etc, in order to properly how these immigrants have integrated in the Spanish Labor Market.

PART I: Analyzing the Panel Data

1.1 Initial analysis

When we merge the two data files: ”Personal” and ”Afiliados”, one containing the personal information of all the individuals, the other containing data about their laboral life. The merge shows that there is data in the ”Personal” database that is not assigned to an id in the ”Afiliacion” database, meaning that there are more individuals in the master database, ”Personal”, than in the other,[1], as we can see in the merge information from stata (the master file was ”Personal”).

Figure 1: **Merge Information**

```
. tab _merge

     Merge  |      Freq.     Percent        Cum.
------------+-----------------------------------
   Personal |    922,925       13.24       13.24
       Both |  6,047,437       86.76      100.00
------------+-----------------------------------
      Total |  6,970,362      100.00
```

As there are more observations than individuals, we can check the number of persons we have in both databases. The results show us that there are 285,207. In order to calculate this, the $tag()$ function in stata is very useful.

[1] There are exactly 922,936 observations that are only present in ”Personal”.

Figure 2: **Merge Information**

```
tag(IDENTP |   Merge
ERS)       |           |     Total
 ----------+-----------+----------
         0 | 5,762,230 | 5,762,230
         1 |   285,207 |   285,207
-----------+-----------+----------
Total      | 6,047,437 | 6,047,437
```

As we can see, only a 4,7% of those observations represent the real number of individuals. This means that, on average, we have 21 observations for each person.

1.2 Analysis of the Labor situation

By dropping all those without any labour data, we get the same number of individuals we got by merging the two data files (logically): 285,207 individuals.

Lets take a look at how the covariates affect the labor situation: gender, age group, and nationality. Before analyzing the data, lets explain the variables. Labor situation is a categorical variable that shows if the worker is employed, self-employed, unemployed or in other situation. The age groups I chose to analyze are three: less than 30 years old, more than 44 years old, and in between. Finally, for the nationalities, we can distinguish the autochthonous (Spanish, for the ease of understanding) and four main groups of immigrants: from the European Union, from Africa, from South America (Latinos) and others.

It is interesting to watch how these covariates have developed throughout the years. For this reason, I have analyzed each one for 2006 and for 2016.

1.2.1 Labor situation

This part showcases the whole sample without covariates.As we can see in 2016 the total number of employed individuals in the labor force has diminished from 145 thousands to 135 (from 75% to 68%). This may probably be because of the 2007 financial recession. The numbers are clear, there are more unemployed workers, and the self-employed numbers have stayed almost the same, albeit a bit smaller.

Figure 3: **Overall Labor Status, 2016**

Labor status	Freq.	Percent	Cum.
Employed	135,782	68.45	68.45
Self-employed	32,170	16.22	84.67
Unemployed	20,933	10.55	95.22
Others	9,483	4.78	100.00
Total	198,368	100.00	

Figure 4: **Overall Labor Status, 2006**

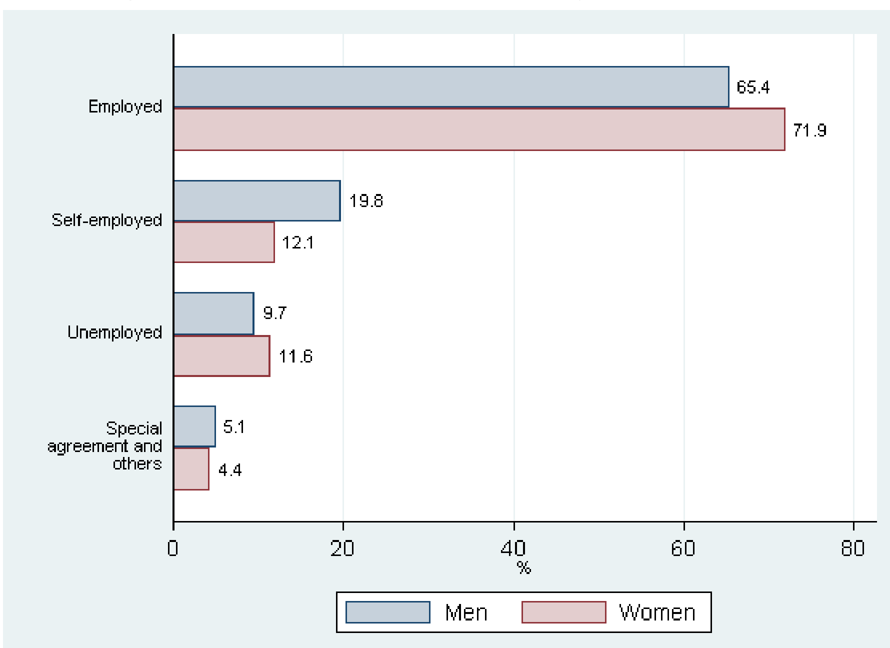
Labor status	Freq.	Percent	Cum.
Employed	143,144	75.62	75.62
Self-employed	31,907	16.86	92.48
Unemployed	12,722	6.72	99.20
Others	1,520	0.80	100.00
Total	189,293	100.00	

1.2.2 Labor situation by Gender

Given the empiricals papers that prove the existance of a gender gap in the labor market regarding wages and employment ()Gneezy, Niederle & Rustichini, 2003), it is interesting to analyze this perspective of the labor situation. In the histogram we can appreciate better the gender gap regarding the labor situation: the women have a higher number of unemployment, and a smaller number for the employed.

If we look at the tables we can appreciate that the difference in the total numbers of workers by gender has diminished from 2006 to 2016, as the difference, within this sample, in 2006 was 32,395 more males in the labor force, and in 2016 is less than 14 thousands. It would be interesting to calculate the change rate of the difference throughout the years to understand the trend. The one clear thing is that, in this aspect, the labor force is more equal than before, with a higher number of women being represented in the sample. The number of males in the labor force has diminished though. If we take a look at the number of self-employed workers, we can see that the males doubles the number of women, and that the number is very stable, as it has not changed almost anything. The number of self-employed women has increased by more than 10% from 2006 to 2016.

Figure 5: Overall Labor situation by Gender (2016)



Source: made with Stata.

Figure 6: **Overall Labor situation by Gender, 2006**

Labor status	Men	Women	Total
Employed	81,894	61,250	143,144
%	73.92	78.03	75.62
Self-employed	21,842	10,065	31,907
%	19.71	12.82	16.86
Unemployed	6,017	6,705	12,722
%	5.43	8.54	6.72
Others	1,041	479	1,520
%	0.94	0.61	0.80
Total	110,794	78,499	189,293
	100.00	100.00	100.00

(Column group header: Sexo, over Men and Women)

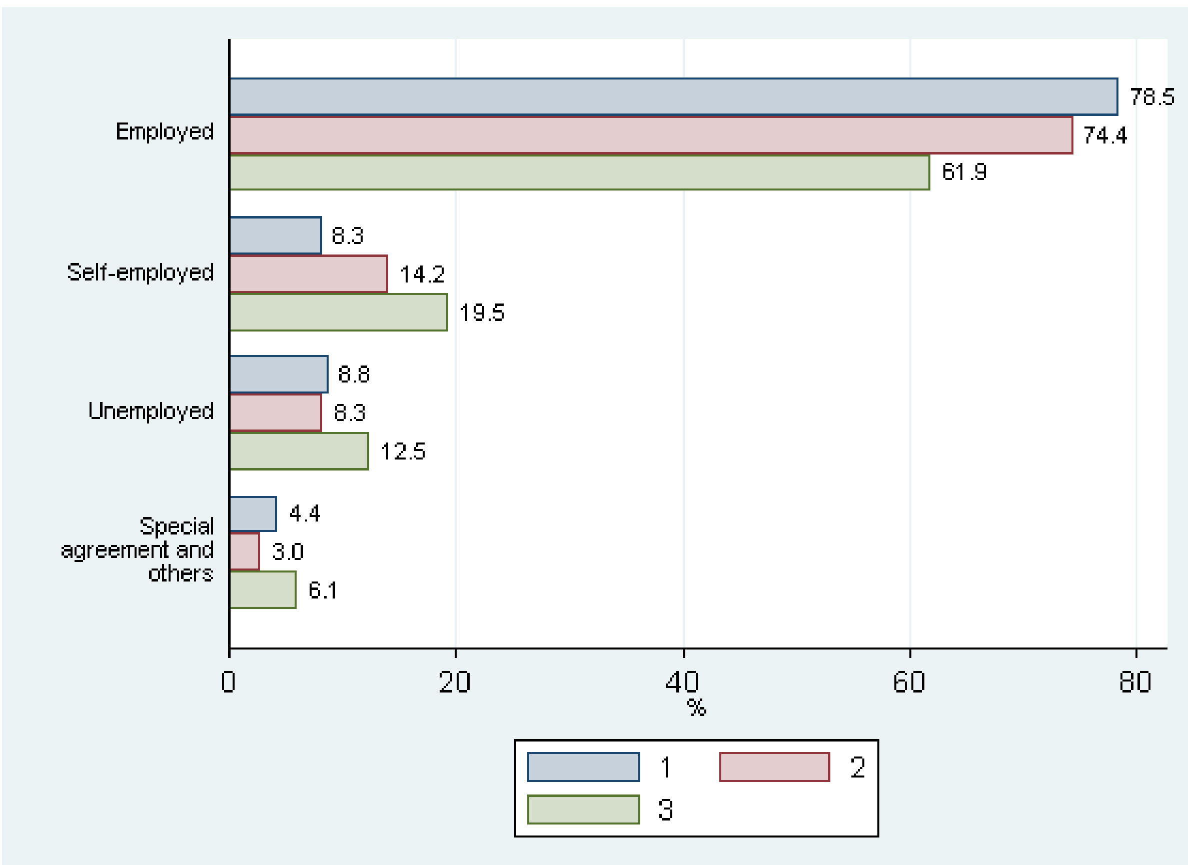
Figure 7: **Overall Labor situation by Gender, 2016**

Labor status	Men	Women	Total
	Sexo		
Employed	69,362	66,420	135,782
%	65.42	71.93	68.45
Self-employed	21,020	11,150	32,170
%	19.82	12.08	16.22
Unemployed	10,255	10,678	20,933
%	9.67	11.56	10.55
Others	5,395	4,088	9,483
%	5.09	4.43	4.78
Total	106,032	92,336	198,368
	100.00	100.00	100.00

1.2.3 Labor situation by Age Groups

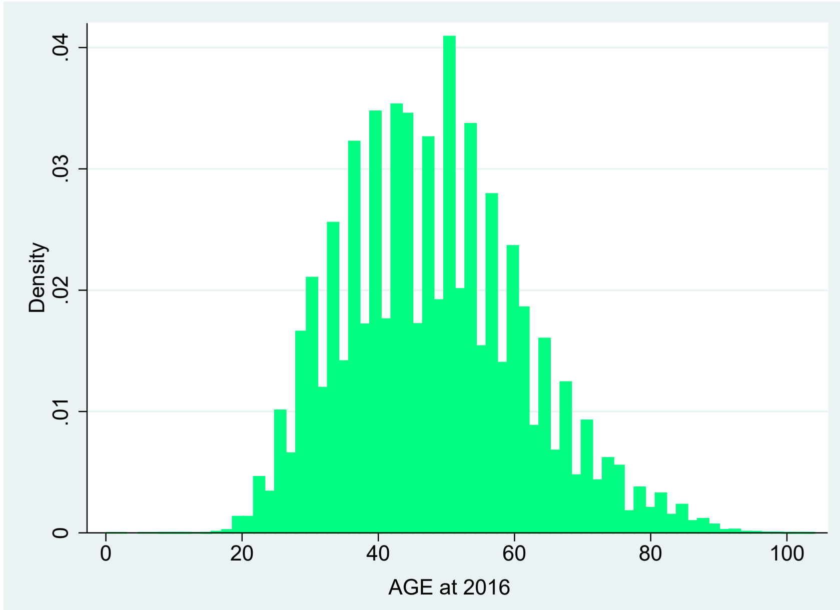
If we take a look at the age groups, we can see that the main part of the workers are in the oldest group considered, that is, older than 44 years. They also represent the highest number of self-employed and the highest proportion of unemployed to total workers ratio, around 20%. We can observe that the labor force in Spain is quite old in 2006, as 67% of the workers (employed, unemployed and self-employed) are older than 44 years old. This changed from 2006 to 2016, as in that last year, the percentage of this age group has diminished to 51.5%, leaving more space for the younger generations. Note that this ”empty space” they left was occupied almost entirely by the youngest age group.

Figure 8: Overall Labor situation by Age Group (on 2016)



Source: made with Stata.

Figure 9: Age Histogram (on 2016)



Source: made with Stata.

Figure 10: **Labor situation by Age Group, 2006**

Labor status	Age Group < 30	(30,44)	>= 44	Total
Employed	2,533	49,583	91,028	143,144
%	94.59	83.41	71.58	75.62
Self-employed	73	6,232	25,602	31,907
%	2.73	10.48	20.13	16.86
Unemployed	72	3,605	9,045	12,722
%	2.69	6.06	7.11	6.72
Others	0	27	1,493	1,520
%	0.00	0.05	1.17	0.80
Total	2,678	59,447	127,168	189,293
	100.00	100.00	100.00	100.00

Figure 11: **Labor situation by Age Group, 2016**

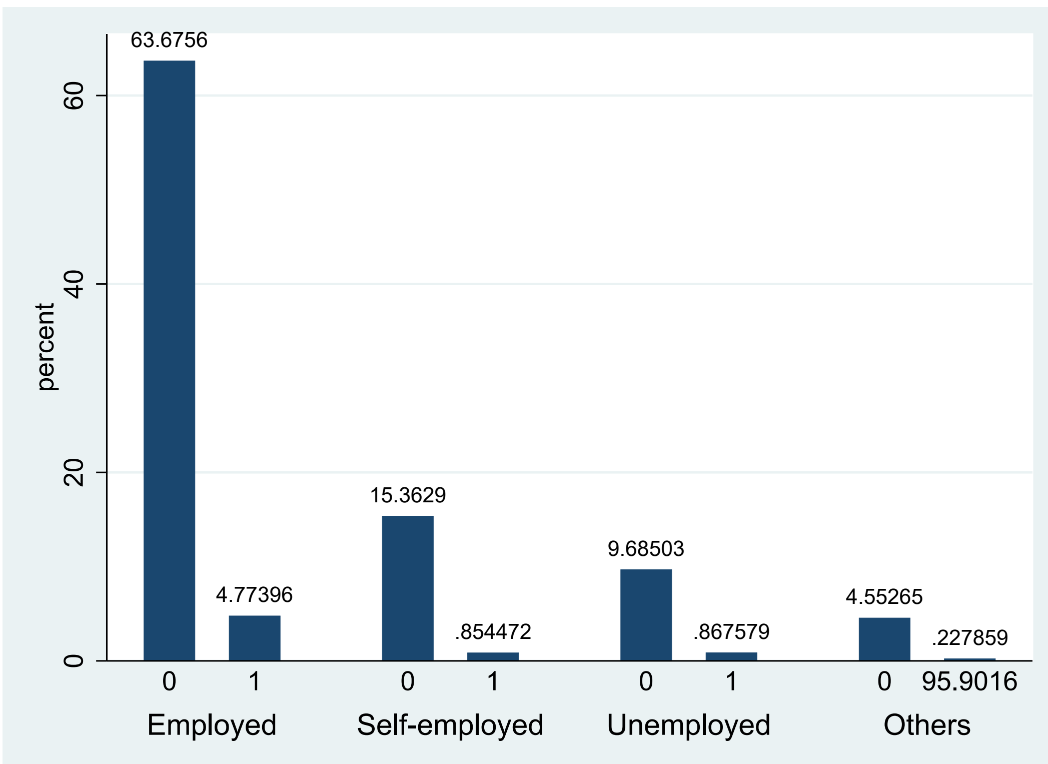
Labor status	Age Group < 30	(30,44)	> 44	Total
Employed	18,785	53,814	63,183	135,782
%	78.47	74.42	61.87	68.45
Self-employed	1,984	10,298	19,888	32,170
%	8.29	14.24	19.47	16.22
Unemployed	2,114	6,031	12,788	20,933
%	8.83	8.34	12.52	10.55
Others	1,055	2,164	6,264	9,483
%	4.41	2.99	6.13	4.78
Total	23,938	72,307	102,123	198,368
	100.00	100.00	100.00	100.00

1.2.4 Labor Situation by inmigrant status

Before analyzing within the immigrants groups, lets see the numbers comparing the whole of them to the "original spanish" people. In the histogram we can appreciate that they actually represent a not very large part of the labor force.

If we take a look at the numbers, we observe that there are less self-employed immigrants than the native people. This is logical, as they have probably arrived recently to the country, and is more difficult to be self-employed. By analyzing how has the percentage of unemployed changed between these two groups from 2006 to 2016, we get that the immigrants suffered more from the crisis, as their unemployment rose from 6% to 12%, whereas for the natives, it rose from 6,7% to 10%.

Figure 12: Overall Labor situation by inmigrant status (on 2016)



Source: made with Stata.

Figure 13: **Labor situation by inmigrant status, 2006**

Labor status	immigrants No	Yes	Total
Employed	134,328	8,816	143,144
%	75.33	80.41	75.62
Self-employed	30,495	1,412	31,907
%	17.10	12.88	16.86
Unemployed	12,056	666	12,722
%	6.76	6.07	6.72
Others	1,450	70	1,520
%	0.81	0.64	0.80
Total	178,329	10,964	189,293
	100.00	100.00	100.00

Figure 14: **Labor situation by inmigrant status, 2016**

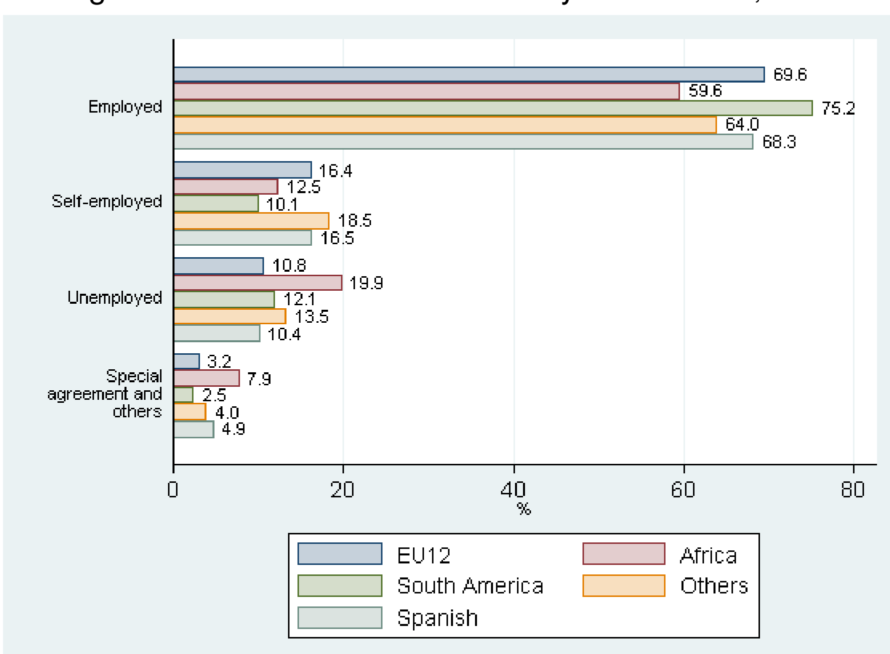
Labor status	immigrants No	Yes	Total
Employed	126,312	9,470	135,782
%	68.27	71.00	68.45
Self-employed	30,475	1,695	32,170
%	16.47	12.71	16.22
Unemployed	19,212	1,721	20,933
%	10.38	12.90	10.55
Others	9,031	452	9,483
%	4.88	3.39	4.78
Total	185,030	13,338	198,368
	100.00	100.00	100.00

1.2.5 Labor Situation by nationalities

Let's now take a look at the labor situation if we consider all nationalities. This is useful to check if there are groups within the immigrants that are more vulnerable. In order to visualize better the differences within all groups lets take a look at the next figure.

South America, has the biggest relative percentage of employed people (also the highest number of workers in absolute values too), and is the most numerous groups within the immigrants. In 2006, that percentage was even higher, reaching a staggering 87% of employed and only a 4,53% of unemployment. If we take a look at the immigrants from the EU, they have the second highest percentage of self-employed, behind the "others" group. This is probably because they are usually better of, coming from a richer country than one in South America or Africa. They also have the lowest (relative) number of unemployed within the immigrants. If we take a look to Africa, they are the most vulnerable group, as their unemployment is the highest by a large margin. The crisis did not affect all groups equally, as the EU12 group only suffered a smaller (relative) increase in the unemployment rate, and Africa went from 80% of employed persons to a 60%. This humongous drop may be because of the stop in the construction sector. Many of these workers take low skill, labor intensive jobs as they are the easiest they can find.

Figure 15: Overall Labor situation by nationalities, 2016



Source: made with Stata.

Figure 16: **Labor Situation by nationalities, 2016**

Labor status	Nationality EU12	Africa	South A.	Others	Spain	Total
Employed	1,244	733	5,959	1,534	126,312	135,782
	69.61	59.64	75.20	63.97	68.27	68.45
Self-employed	293	154	804	444	30,475	32,170
	16.40	12.53	10.15	18.52	16.47	16.22
Unemployed	193	245	960	323	19,212	20,933
	10.80	19.93	12.12	13.47	10.38	10.55
Others	57	97	201	97	9,031	9,483
	3.19	7.89	2.54	4.05	4.88	4.78
Total	1,787	1,229	7,924	2,398	185,030	198,368
	100.00	100.00	100.00	100.00	100.00	100.00

Figure 17: *

Labor Situation by nationalities, 2006

	Nationality					
Labor status	EU12	Africa	South A.	Others	Spain	Total
Employed	1,247	994	4,555	2,020	134,328	143,144
	73.57	80.75	87.06	71.99	75.33	75.62
Self-employed	314	133	429	536	30,495	31,907
	18.53	10.80	8.20	19.10	17.10	16.86
Unemployed	128	94	237	207	12,056	12,722
	7.55	7.64	4.53	7.38	6.76	6.72
Others	6	10	11	43	1,450	1,520
	0.35	0.81	0.21	1.53	0.81	0.80
Total	1,695	1,231	5,232	2,806	178,329	189,293
	100.00	100.00	100.00	100.00	100.00	100.00

1.3 Analysis of employed individuals

Lets take a look at the contract types of these individuals, by having the same covariates. In Spain, we can distinguish two distinct labor forces, one stable, well-paid and with indefinite contracts; and one temporal, worse-paid, no bonuses, etc.

First of all, I will display the general distribution of the contract types. In this sample there are five main categories: indefinite partial time, indefinite full time, temporal partial and full time; and unknown. As we can appreciate in the following table, almost all of the contracts are indefinite and full time, although the percentage of temporal contracts is a serious issue, as it is a very high value (26,1%).

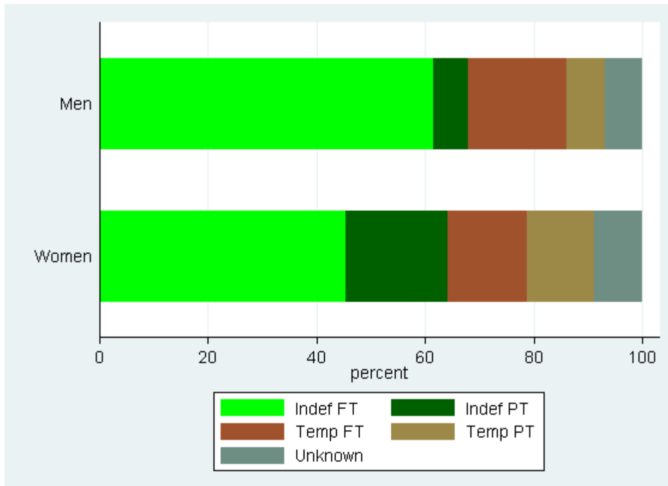
Figure 18: **Overall statistics of the type of contracts**

Type of contract	Freq.	Percent	Cum.
Indef FT	72,790	53.61	53.61
Indef PT	16,902	12.45	66.06
Temp FT	22,285	16.41	82.47
Temp PT	13,169	9.70	92.17
Unknown	10,636	7.83	100.00
Total	135,782	100.00	

If we consider the gender as a covariate, we can see that there is a high imbalance between as the males have a higher proportion of indefinite contracts, and within them, a higher number of full-time ones, and a lower for the partial time. The cause of this disproportion could be attributed to the fact that the women are more involved with the children and the house chores than the men, and this forces them to take partial time contracts.

The same happens with the temporal contracts, where women display a higher relative number, influenced also by the fact that the pregnancy is not very supported in Spain, and thus women are more prone to suffer from sex discrimination than in others first world countries.

Figure 19: Type of Contracts by Gender, 2016



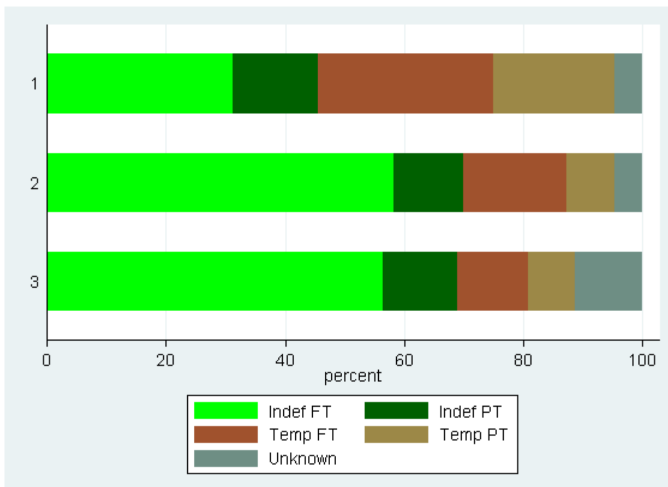
Source: made with Stata.

Figure 20: **Type of Contracts by Gender, 2016**

Type of contract	Sexo: Men	Sexo: Women	Total
Indef FT	42,670	30,120	72,790
	61.52	45.35	53.61
Indef PT	4,452	12,450	16,902
	6.42	18.74	12.45
Temp FT	12,544	9,741	22,285
	18.08	14.67	16.41
Temp PT	4,942	8,227	13,169
	7.12	12.39	9.70
Unknown	4,754	5,882	10,636
	6.85	8.86	7.83
Total	69,362	66,420	135,782
	100.00	100.00	100.00

If we consider the age as the covariate, we can observe that the highest percentage of full time, indefinite contracts, is held by the middle age group. The youngest group have the highest proportion of temporal and partial time contracts, which showcases how adverse is the labor market for young people in Spain. The oldest group do not have the highest proportional number of indefinite contracts, but they do have the lowest number of temporal contracts, which make sense, as these jobs are usually the most precarious ones, and most of them come from the tourism service sector, where the nature of the young often requires younger persons.

Figure 21: Type of Contracts by Age, 2016



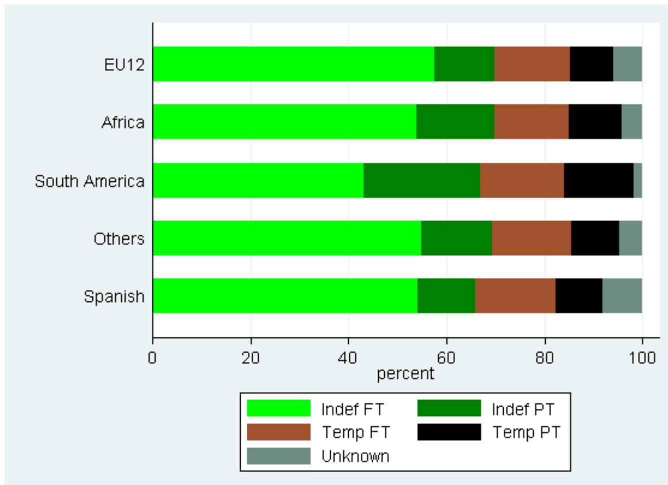
Source: made with Stata.

Figure 22: **Type of Contracts by Age, 2016**

Type of contract	Age Group < 30	(30,44)	> 44	Total
Indef FT	5,858	31,281	35,651	72,790
	31.18	58.13	56.42	53.61
Indef PT	2,695	6,283	7,924	16,902
	14.35	11.68	12.54	12.45
Temp FT	5,506	9,370	7,409	22,285
	29.31	17.41	11.73	16.41
Temp PT	3,828	4,384	4,957	13,169
	20.38	8.15	7.85	9.70
Unknown	898	2,496	7,242	10,636
	4.78	4.64	11.46	7.83
Total	18,785	53,814	63,183	135,782
	100.00	100.00	100.00	100.00

Taking the nationality groups as covariates let us analyze much better the laboral situation within the immigrants. We can observe that all of the immigrants showcase a higher proportion of temporal jobs than the average Spaniard, with Africa being the worst off, with more than 30%, with the others two distinguishable groups being very close in between them. This is also displayed in the percentage of the highest group of immigrants, Africa, where they have the lowest percentage of indefinite full time contracts.

Figure 23: Type of Contracts by Nationality, 2016



Source: made with Stata.

Figure 24: **Type of Contracts by Nationality, 2016**

Type of contract	Nationality España	EU12	Africa	South Ame	Others	Total
Indef FT	716	395	2,567	841	68,271	72,790
	57.56	53.89	43.08	54.82	54.05	53.61
Indef PT	153	116	1,419	220	14,994	16,902
	12.30	15.83	23.81	14.34	11.87	12.45
Temp FT	192	111	1,019	250	20,713	22,285
	15.43	15.14	17.10	16.30	16.40	16.41
Temp PT	108	79	845	150	11,987	13,169
	8.68	10.78	14.18	9.78	9.49	9.70
Unknown	75	32	109	73	10,347	10,636
	6.03	4.37	1.83	4.76	8.19	7.83
Total	1,244	733	5,959	1,534	126,312	135,782
	100.00	100.00	100.00	100.00	100.00	100.00

Another way to properly understand the labor market, would be by distinguishing the average of worked hours per day with these same covariates. If we take a look at the overall statistics, we observe that the standard deviation among the full time contracts is very low, they all have around 7,8 hours per day of worked hours on average. The difference is visibly higher between the partial time contracts, with the temporal one being almost one worked hour behind the other. If we consider the standard deviation inside the contracts, we can see that the full times ones has less than half of the standard deviation of the temporal ones.

Figure 25: **Overall statistics of Worked Hours, 2016**

Type of contract	Summary of Worked Hours Mean	Std. Dev.	Freq.
Indef FT	7.8081598	.85415417	72,790
Indef PT	4.8280895	2.1440006	16,902
Temp FT	7.9682911	.32342251	22,285
Temp PT	3.9705071	2.1347948	13,169
Unknown	7.8700188	.69336688	10,636
Total	7.0961307	1.8685441	135,782

If we analyze the average worked hours by the gender, we see further proof of a gender gap in the labor market: the standard deviation of the worked hours (which probably means that the contracts are not as stable as others with low standard deviation), is higher for the women in all of the contract types, being more than double of the men in some of them. The difference with the average values is very high in the partial time contracts, specially in the partial time temporal contract, where men have as average 3,5 worked hours, and women 4,24.

As we can see in the next simple regression where I regress the worked hours by gender, the women work on average, less hours. This is because of the higher proportion of temporal contracts. It would be interesting to make a multilogit to properly asses the effect of gender for each type of contract, although the tables showcases these differences pretty well also.

Table 1: Regressing Worked Hours by Gender

	(1)
	hworked
Men	0
	(.)
Women	-0.447***
	(-345.95)
Constant	7.601***
	(8985.79)
Observations	6047437
Adjusted R^2	0.019

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 26: **Worked hours by Gender, 2016**

Type of contract	Sexo M	F	Total
Indef FT	7.9538107	7.601821	7.8081598
Std. Dev	.46588903	1.1760351	.85415417
Frequency	42670	30120	72790
Indef PT	4.3375112	5.0035155	4.8280895
Std. Dev	1.9649798	2.1779043	2.1440006
Frequency	4452	12450	16902
Temp FT	7.9898508	7.9405277	7.9682911
Std. Dev	.19397923	.43530622	.32342251
Frequency	12544	9741	22285
Temp PT	3.5120437	4.2459084	3.9705071
Std. Dev	1.9590368	2.1881401	2.1347948
Frequency	4942	8227	13169
Unknown	7.9203601	7.8293315	7.8700188
Std. Dev	.58032138	.7703914	.69336688
Frequency	4754	5882	10636
Total	7.4094505	6.7689327	7.0961307
Std. Dev	1.6228594	2.0439447	1.8685441
Frequency	69362	66420	135782

When we take into consideration the age, we can see that in the full time indefinite contracts there is practically no difference between the age groups. This is also true for the indefinite partial time contracts and the full time temporal contracts, both on average and in volatility. The contract with the highest differences between the groups is the most precarious one: the partial time, temporal contracts. Overall age do not affect neither the average value nor the volatility of the worked hours. I was surprised by this, so I made a very simple regression to see if the age groups were really not relevant. As we can see in the regression, the results show that they are indeed relevant variables. The benchmark is the youngest group and thus, we see that the middle group works less than, and the oldest more, maybe because of them having higher ranking in their jobs and having to spend more time.

Table 2: Regressing Worked Hours by Age groups

	(1)
	hworked
< 30	0
	(.)
30 - 43	0.680***
	(255.74)
≥ 44	1.092***
	(428.33)
Constant	6.526***
	(2699.88)
Observations	6047437
Adjusted R^2	0.037

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 27: **Worked hours by Age, 2016**

Means, Standard Deviations and Frequencies of hworked

Type of contract	Age Group <30	(30,44)	>44	Total
Indef FT	7.7717788	7.7647901	7.8521913	7.8081598
Std. Dev	1.0657358	.86869196	.79818534	.85415417
Frequency	5858	31281	35651	72790
Indef PT	4.477741	4.890982	4.8973771	4.8280895
Std. Dev	1.9555311	2.0063614	2.2943101	2.1440006
Frequency	2695	6283	7924	16902
Temp FT	7.9836527	7.9495351	7.9805955	7.9682911
Std. Dev	.27654142	.39783644	.238497	.32342251
Frequency	5506	9370	7409	22285
Temp PT	3.889814	4.2749179	3.763599	3.9705071
Std. Dev	1.9009519	2.0365319	2.3495976	2.1347948
Frequency	3828	4384	4957	13169
Unknown	7.8627082	7.7434295	7.9145551	7.8700188
Std. Dev	.74141162	.92529938	.57920598	.69336688
Frequency	898	2496	7242	10636
Total	6.574581	7.1761331	7.1830538	7.0961307
Std. Dev	2.2125823	1.7115889	1.8597995	1.8685441
Frequency	18785	53814	63183	135782

Finally, if we distinguish between the different nationalities, we observe that for the indefinite full time contracts, the average values and the volatilities are very close among the groups, with the natives working the least time of them. This is not true for the partial time contracts in the indefinite ones, where the average values are much higher for the african and the latinos immigrants. This also happens in the partial time, temporal contracts. As I did earlier with the age, given that I will be analyzing the immigration in the second part of the assignment, I also did a regression. As I am not distinguishing by type of contracts, the coefficients do not make lot of sense for the African coefficient for example, as they have a very high percentage of temporal jobs and thus they will work more hours. But the regression is useful for one thing, it showcases that, on average of all the contract types, the immigrants usually work more hours.

Table 3: Regressing Worked Hours by Nationalities

	(1)
	hworked
España	0
	(.)
EU12	0.176***
	(16.72)
Africa	-0.235***
	(-30.10)
South America	0.163***
	(19.41)
Others	0.0202**
	(2.94)
Constant	7.393***
	(1080.44)
Observations	6047437
Adjusted R^2	0.001

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 28: **Worked hours by Nationality, 2016**

Means, Standard Deviations and Frequencies of hworked

Type of contract		Nationality EU12	Africa	Latinos	Others	Spain	Total
Indef FT		7.8144916	7.9017924	7.8839829	7.918887	7.8033367	7.8081598
Std. Dev		.7873041	.62271049	.66972189	.5637494	.86483085	.85415417
Frequency		716	395	2567	841	68271	72790
Indef PT		4.9006013	5.1451724	5.4700493	4.9892727	4.7617778	4.8280895
Std. Dev		2.180369	2.4449712	2.4223645	2.1263772	2.1028655	2.1440006
Frequency		153	116	1419	220	14994	16902
Temp FT		7.97375	8	7.9763847	8	7.9672897	7.9682911
Std. Dev		.2657923	0	.27771717	0	.32875053	.32342251
Frequency		192	111	1019	250	20713	22285
Temp PT		4.3493333	4.417519	4.8191337	4.08224	3.9029275	3.9705071
Std. Dev		2.1792584	2.3145684	2.4030301	2.1612777	2.0992196	2.1347948
Frequency		108	79	845	150	11987	13169
Unknown		7.9432533	8	7.8676697	7.8350685	7.8693573	7.8700188
Std. Dev		.33491762	0	.62216591	.70279522	.69699803	.69336688
Frequency		75	32	109	73	10347	10636
Total		7.1876206	7.1091842	6.8900594	7.1328031	7.1044303	7.0961307
Std. Dev		1.750507	1.8989286	2.0399445	1.8309407	1.8609204	1.8685441
Frequency		1244	733	5959	1534	126312	135782

1.4 Panel data Observations

Before further analysis, we will be manipulating once more the data in order to do a better study. The created panel, is a database that contains two observations per year per individual: the first day of May, and the first day of October, before and after the summer, which makes sense, as in Spain there is a high increase in employment during those summer months. The considered time period is made of eleven years, from 2006 to 2016. This way we should have 22 observations for each individual. I also delete those workers older than 45 years, this way the heterogeneity among our sample of individuals is decreased. Now we have a database where there are 108,564 individuals, from the 285,207 we had before. These individuals will have 22 observations each, making a total of 2,388,408. Lets check if this is true:

Figure 29: **Panel Data Observations**

```
Contains data from data_2.dta
obs:     2,388,408
vars:           45                         12 Apr 2020 03:03
size:   229,287,168
```

1.5 Panel data variables

These are some of the most important variables I will be using in this assignment.

Table 4: Most important variables of the dataset

Variable	Observations	Mean	Comment
IDENTPERS	-	-	Id of the person
t	2388408	11.5	Time variable
FALTA	1706897	17833.62	Date of entering labor force
FBAJA	1706897	19884.45	Date of exiting labor force
dbirth	2388408	7.955,599	Date of birth
nacionalidad	-	-	Nacionality
province	2388408	27,463	Province of Birth
gender	2388408	1,47726	Self explanatory
situlab	2388408	1,80600	Labor situation
tipo	1334554	1,92168	Type of contract
grupcot	1334554	5,98864	Quote group
age_oct16	2,388,408	34.51206	Age at October 2016
hworked	1,334,554	7.237032	Average of worked hours
hourly_wage	1,325,198	11.65185	Wage per hour worked
monthly_wages	2,388,408	1574.498	Wage per month worked

PART II: Analysis of Immigrants in the Spanish Labor Market

2.1 Data Description

Before doing a deeper analysis of the data, lets study the distribution of the individuals with respect to the covariates:

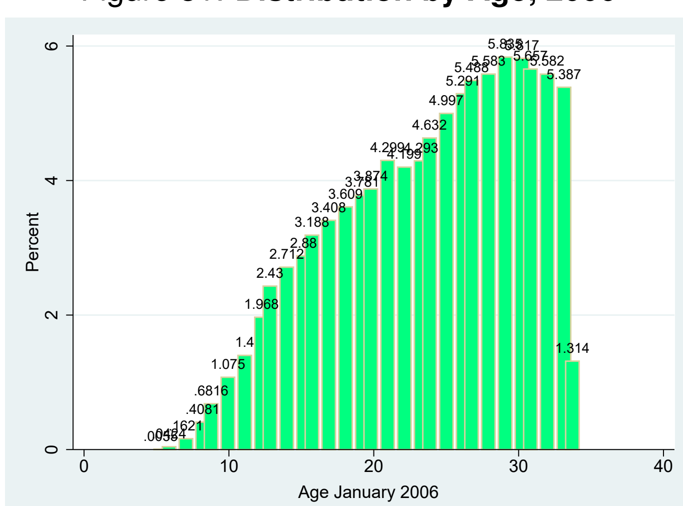
As we can observe, the genders are very much equally distributed, as the percentage is around 50% for both of them, with 56,751 men and 51,813 women.

Figure 30: **Distribution by Gender, 2006**

Gender	Freq.	Percent	Cum.
Male	56,751	52.27	52.27
Female	51,813	47.73	100.00
Total	108,564	100.00	

Regarding the age, the pyramid is much more concentrated to the older individuals, showcasing that the labor force in Spain has a bigger percentage of older individuals.

Figure 31: **Distribution by Age, 2006**



Source: made with Stata.

If we discriminate by nationality, we see that 93% of the individuals are Spanish, logically. So, we have a percentage of 7,04% of immigrants. The biggest group is the South American, whom we share a very similar culture.

Figure 32: **Distribution by Nationality, 2006**

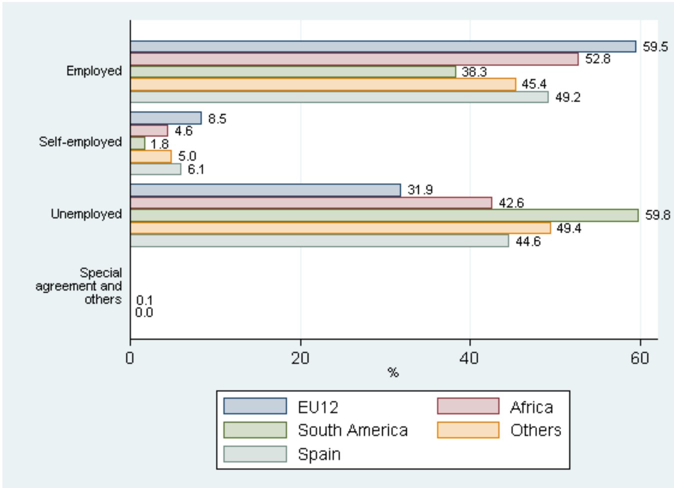
Nationality	Freq.	Percent	Cum.
EU12	749	0.69	0.69
Africa	591	0.54	1.23
Latino	4,753	4.38	5.61
Other	1,545	1.42	7.04
Spain	100,926	92.96	100.00
Total	108,564	100.00	

For further improvement of the analysis, I will be analyzing the labor market conditions of the individuals in my sample: the labor situation, the type of contracts, the worked hours, the monthly and the hourly wages, for the first period (May of 2006) and for the last (October of 2016).

In 2006, we can observe that, by eliminating those individuals older than 44 years old, the unemployment has risen considerably to a 45%. This showcases the big problem of young unemployment in Spain, as almost all of the unemployed individuals are within the first two younger age groups (as stata shows). Within the inmigrant groups, those that come from the European Union has the highest rate of employment, superior to the Spanish, even. Africa comes second, and then the Spanish. The South American immigrants are the ones that suffer most from the unemployment: 59,84%.

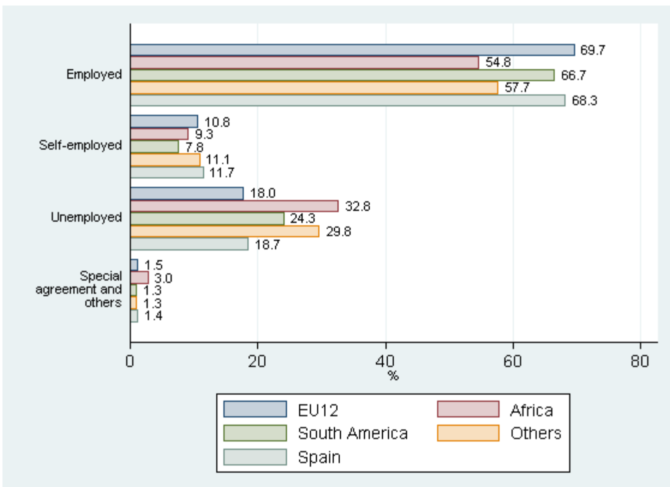
If we take a look at the 2016 graph, we can appreciate that the employment has risen considerably, this is probably linked to the fact that the individuals are 10 years older and thus, the probability of being unemployed has diminished: the overall unemployment rate is 19%, whereas before it was close to 45%. Within the groups, the Latinos experienced the biggest drop in the unemployment rate, as they went from almost 60% to 7%! The self-employed individuals have doubled since then, but not equally for all groups, as the Spanish and Latino groups comprises most of them.

Figure 33: **Labor Situation, 2006**



Source: made with Stata.

Figure 34: **Labor Situation, 2016**



Source: made with Stata.

Figure 35: **Labor Situation by Nationality, May 2006**

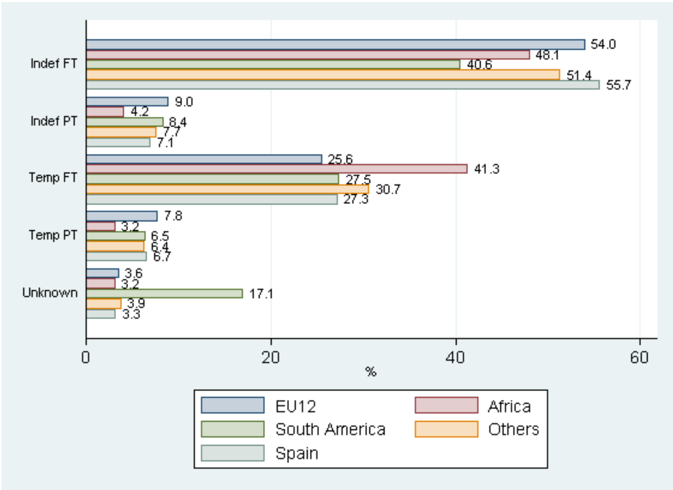
	Nationality					
Labor status	EU12	Africa	Latinos	Others	Spain	Total
Employed	446	312	1,822	701	49,680	52,961
	59.55	52.79	38.33	45.37	49.22	48.78
Self-employed	64	27	87	78	6,185	6,441
	8.54	4.57	1.83	5.05	6.13	5.93
Unemployed	239	252	2,844	764	45,040	49,139
	31.91	42.64	59.84	49.45	44.63	45.26
Others	0	0	0	2	21	23
	0.00	0.00	0.00	0.13	0.02	0.02
Total	749	591	4,753	1,545	100,926	108,564
	100.00	100.00	100.00	100.00	100.00	100.00

Figure 36: **Labor Situation by Nationality, October 2016**

Labor status	EU12	Africa	Latinos	Others	Spain	Total
	Nationality					
Employed	522	324	3,168	892	68,926	73,832
	69.69	54.82	66.65	57.73	68.29	68.01
Self-employed	81	55	370	172	11,769	12,447
	10.81	9.31	7.78	11.13	11.66	11.47
Unemployed	135	194	1,153	461	18,834	20,777
	18.02	32.83	24.26	29.84	18.66	19.14
Others	11	18	62	20	1,397	1,508
	1.47	3.05	1.30	1.29	1.38	1.39
Total	749	591	4,753	1,545	100,926	108,564
	100.00	100.00	100.00	100.00	100.00	100.00

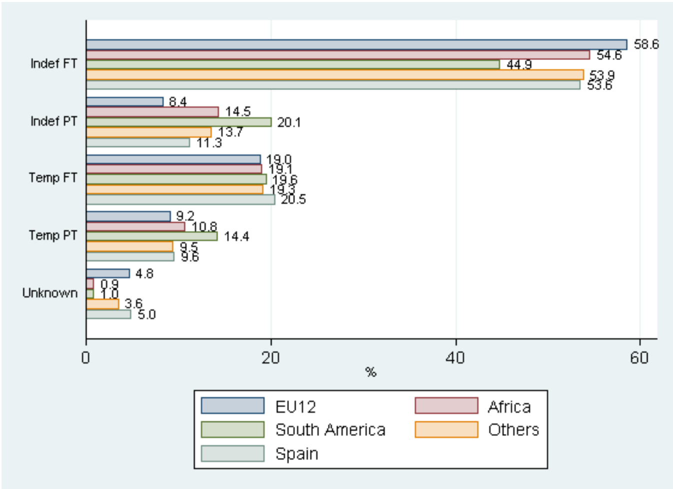
Analyzing by type of contract we find that we have less contracts than individuals. This is because of the unemployed workers inside the sample. That's why in the first period the number of contracts is much less than in the last one, as the individuals grow older have a higher chance of being employed. We get almost the same conclusions that we drew from the data in the first part, the native and the Eu12 groups have the best contracts: indefinite, full time jobs, whereas the African, and specially the latinos have the worst distribution, as they have few indefinite contracts, and many temporal ones.

Figure 37: **Type of Contracts, 2006**



Source: made with Stata.

Figure 38: **Type of Contracts, 2016**



Source: made with Stata.

In the last period, the situation has worsened for the Latino and African groups for some type of contracts, as the temporal contracts they had (most of them full time) have diversified into the partial time temporal contracts, having the full time about 20% of the percentage, and the partial time between 10-15%. If we analyze the indefinite one, all of the groups has increased their share among the two types, or has stayed more or less the same (like Spain). If we analyze the overall situation, one there are more indefinite contracts, but also more partial time temporal jobs.

Figure 39: **Type of Contracts by Nationality, October 2006**

Type of contract	Nationality EU12	Africa	Latinos	Others	España	Total
Indef FT	241	150	739	360	27,647	29,137
	54.04	48.08	40.56	51.36	55.65	55.02
Indef PT	40	13	153	54	3,525	3,785
	8.97	4.17	8.40	7.70	7.10	7.15
Temp FT	114	129	501	215	13,543	14,502
	25.56	41.35	27.50	30.67	27.26	27.38
Temp PT	35	10	118	45	3,324	3,532
	7.85	3.21	6.48	6.42	6.69	6.67
Unknown	16	10	311	27	1,641	2,005
	3.59	3.21	17.07	3.85	3.30	3.79
Total	446	312	1,822	701	49,680	52,961
	100.00	100.00	100.00	100.00	100.00	100.00

Figure 40: **Type of Contracts by Nationality, October 2016**

Type of contract	Nationality: EU12	Africa	Latinos	Others	España	Total
Indef FT	306	177	1,423	481	36,922	39,309
	58.62	54.63	44.92	53.92	53.57	53.24
Indef PT	44	47	638	122	7,821	8,672
	8.43	14.51	20.14	13.68	11.35	11.75
Temp FT	99	62	621	172	14,120	15,074
	18.97	19.14	19.60	19.28	20.49	20.42
Temp PT	48	35	455	85	6,628	7,251
	9.20	10.80	14.36	9.53	9.62	9.82
Unknown	25	3	31	32	3,435	3,526
	4.79	0.93	0.98	3.59	4.98	4.78
Total	522	324	3,168	892	68,926	73,832
	100.00	100.00	100.00	100.00	100.00	100.00

Lets analyze by the worked hours. In 2006 almost all of the inmigrant groups share the same amount of worked hours, maybe Africa is a bit above the rest. The standard deviation, that is, the volatility, is practically the same for all groups. In 2016, this does not hold, as the volatility is now higher overall and more for the Latinos and African individuals. Regarding the average value, all of them now work less than before, with the latinos being the ones that suffered the biggest decrease.

Figure 41: **Worked Hours by Nationality, October 2006**

	Summary of Worked Hours		
Nationality	Mean	Std. Dev.	Freq.
EU12	7.4079641	1.4291859	446
Africa	7.6978974	1.1668299	312
South Ame	7.4535368	1.4193309	1,822
Others	7.4653238	1.3995986	701
Spain	7.4352229	1.4956772	49,680
Total	7.4375692	1.4897004	52,961

Figure 42: **Worked Hours by Nationality, October 2016**

	Summary of Worked Hours		
Nationality	Mean	Std. Dev.	Freq.
EU12	7.2349885	1.6233286	522
Africa	7.0231605	1.8924414	324
South Ame	6.8486313	1.9834179	3,168
Others	7.1184395	1.8001005	892
Spain	7.1171101	1.748809	68,926
Total	7.1060273	1.7607762	73,832

Analyzing the monthly wages we find that the natives and the European groups are among the highest monthly wages, whereas the African and the Latinos the lowest. In 2016, wages have increased, but now the difference between the highest paid and the lowest id bigger in absolute values. Lets take a look at the hourly wages.

Figure 43: **Monthly Wages by Nationality, October 2006**

Nationality	Obs	Mean	Std. Dev.	Min	Max
EU12	749	1337.241	0	1337.241	1337.241
Africa	591	1149.665	0	1149.665	1149.665
Latinos	4,753	1013.867	0	1013.867	1013.867
Others	1,545	1340.5	0	1340.5	1340.5
Spain	100,926	1311.273	0	1311.273	1311.273

Figure 44: **Monthly Wages by Nationality, October 2016**

Nationality	Obs	Mean	Std. Dev.	Min	Max
EU12	749	1720.583	0	1720.583	1720.583
Africa	591	1384.914	0	1384.914	1384.914
Latinos	4,753	1332.977	0	1332.977	1332.977
Others	1,545	1731.556	0	1731.556	1731.556
Spain	100,926	1714.917	0	1714.917	1714.917

By taking a look at the hourly wages we can appreciate that the Latinos have the highest rates, while having the least monthly wages. Once again, the Spaniards and the Europeans are among the median or the highest percentile. The Africans have the lowest hourly wages.

Figure 45: **Hourly Wages by Nationality, October 2006**

Nationality	Obs	Mean	Std. Dev.	Min	Max
EU12	749	8.13084	0	8.13084	8.13084
Africa	591	7.036799	0	7.036799	7.036799
Latinos	4,753	10.44046	0	10.44046	10.44046
Others	1,545	8.236124	0	8.236124	8.236124
Spain	100,926	9.309595	0	9.309595	9.309595

Figure 46: **Hourly Wages by Nationality, October 2016**

Nationality	Obs	Mean	Std. Dev.	Min	Max
EU12	749	12.09313	0	12.09313	12.09313
Africa	591	9.248952	0	9.248952	9.248952
Latinos	4,753	9.583117	0	9.583117	9.583117
Others	1,545	11.60483	0	11.60483	11.60483
Spain	100,926	12.47974	0	12.47974	12.47974

Furthering the analysis of hourly wages and the Latino community, lets take a look at the following graphs, one is for the monthly wages between natives and Latinos, the other for the hourly wages. We can see clearly that, despite the natives having a much higher monthly wages, the Latinos have higher hourly wages.[2] This is probably cause of the higher rate of temporal contracts of the Latinos, with jobs with a higher payment per hour, albeit less hours worked. In the following table, we can appreciate the results of the regression of the hourly wages, which has as independent variables two categoricals, the nationality and the contract types, and one continuous one, the worked hours:

$$hourly\_wage = \beta_0 + \beta_1 Nationality + \beta_2 Tipo + \beta_3 workedhours. \quad (1)$$

As we can see in the results, being a South American individuals has a positive coefficient, that is, the hourly wages are increased. This would need further investigation as we can see that the temporal contracts have a negative coefficient, which makes invalid my previous assumption. One thing is clear, the Latinos have the highest hourly wages among all the groups.

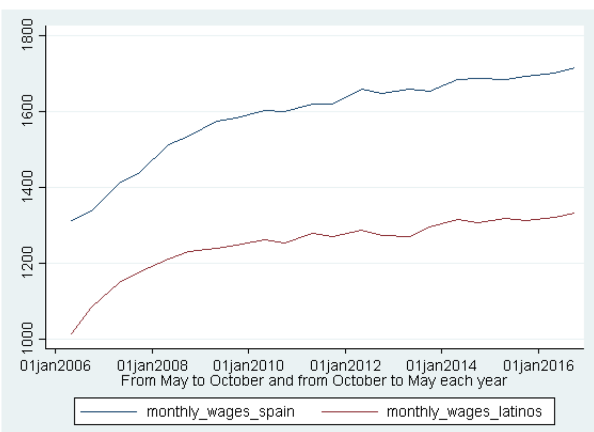


Figure 47: Monthly Wage

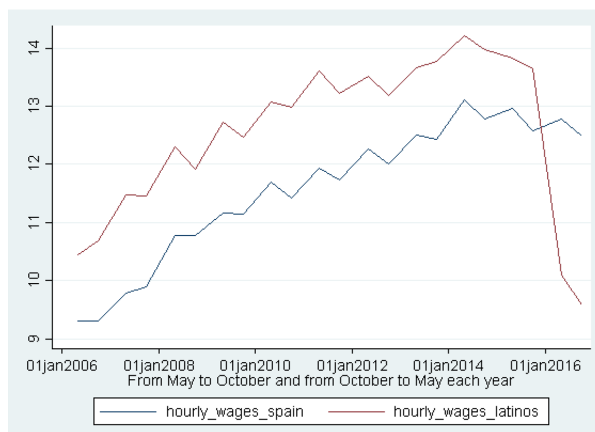


Figure 48: Hourly Wages

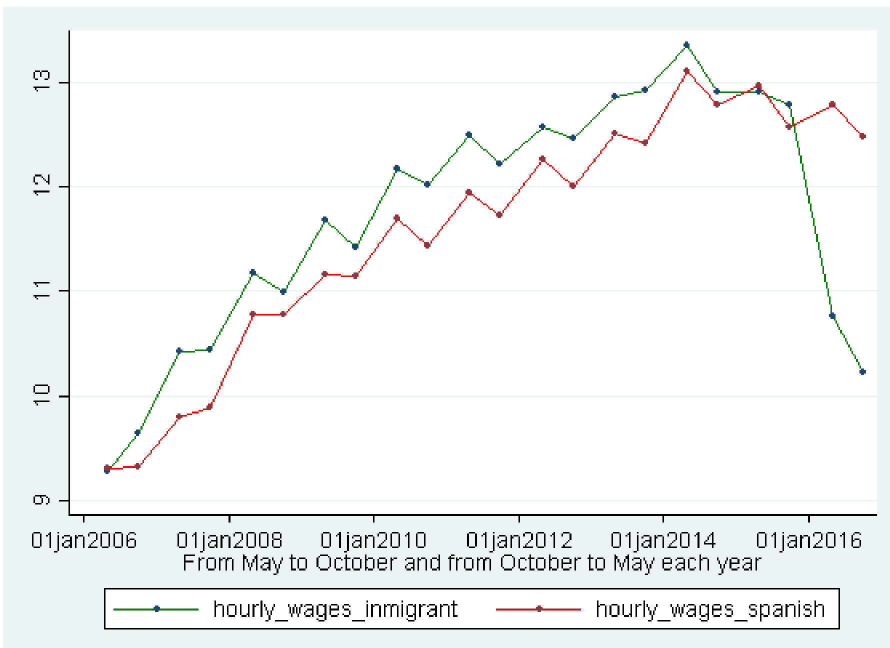
[2] We can also appreciate a very high drop in the hourly wages of the Latinos toward the end of the time period.

2.2 Visual representation

Before making the regressions to finally delve deep into the analysis of the data, I will make a few graphs that will showcase the evolution of the hourly and the monthly wages between the immigrants, and the natives people.

In this first graph, we can see that both of the variables follow the same trend for the most part of the time period. Towards the end, the hourly wages of the immigrants descend abruptly, leaving the natives at the top. I assume that the reason for the hourly wages of the immigrants to be superior than the natives ones, may lay in the higher percentage of temporal jobs, or in the rate of unemployment.

Figure 49: **Hourly Wages**



Source: made with Stata.

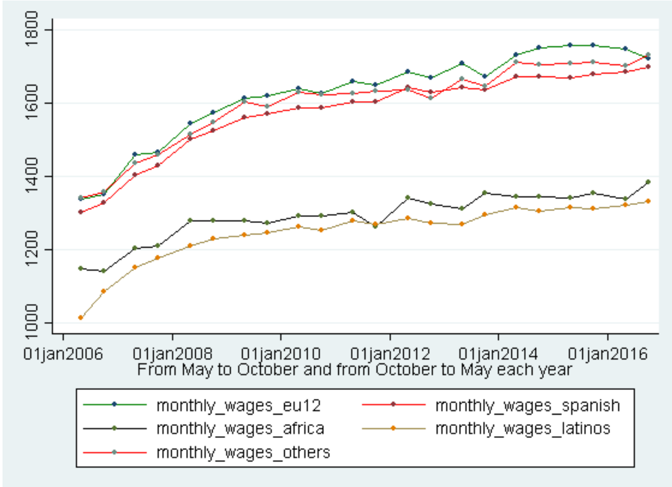
In the second, the monthly wages are displayed within the groups, for a better understanding of the wages. Here we can see that the highest monthly rates corresponds to the Europeans, the Spanish and others, whereas the Latinos and the African individuals are way below, although they have the biggest share in the inmigrant count.

Table 5: Regressing Hourly Wages

	(1)
	hourly_wage
EU12	-0.369
	(-0.34)
Africa	-2.771*
	(-2.06)
South America	1.070*
	(2.14)
Others	-0.0964
	(-0.11)
Spain	0
	(.)
Indef FT	0
	(.)
Indef PT	-33.20***
	(-80.64)
Temp FT	-2.106***
	(-8.66)
Temp PT	-42.03***
	(-87.95)
Unknown	-1.129*
	(-2.39)
hworked	-10.65***
	(-124.59)
Constant	95.60***
	(141.58)
Observations	1325198
Adjusted R^2	0.012

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

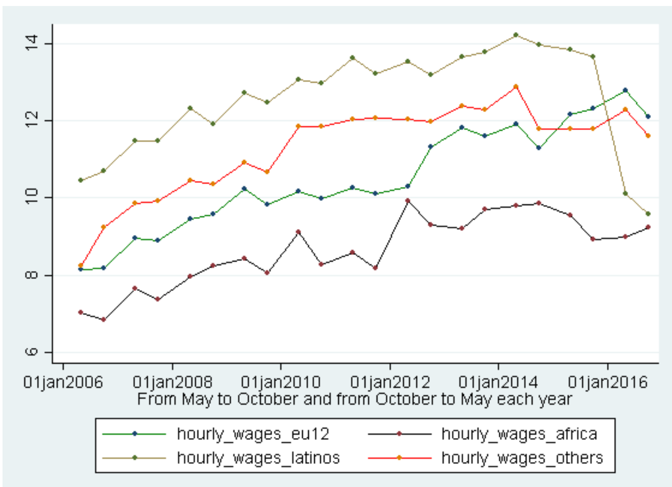
Figure 50: **Monthly Wages**



Source: made with Stata.

Lets analyze only the differences within the inmigrant groups. If we take a look at the evolution of the hourly wages, we can see that the latinos are above the rest, as we saw before, except towards the end, where the wages plummet. They all follow the same trend, more or less, with the African hourly wages being the smallest one throughout the years, although in the end, the wage is very similar to the Latinos one.

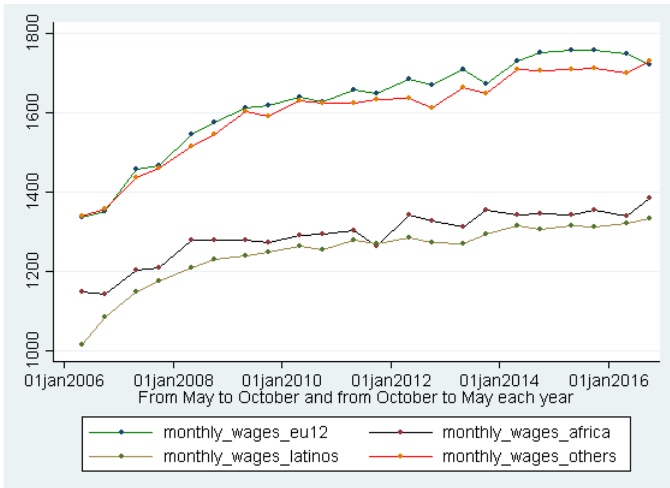
Figure 51: **Hourly Wages within inmigrant groups**



Source: made with Stata.

Analyzing the trend of the monthly wages within inmigrant groups, we find that they are very similar between them, in pairs. The first one, the "others" group and the Europeans, share a very high monthly wages, if we compare them to the other pair, the African and Latino one.

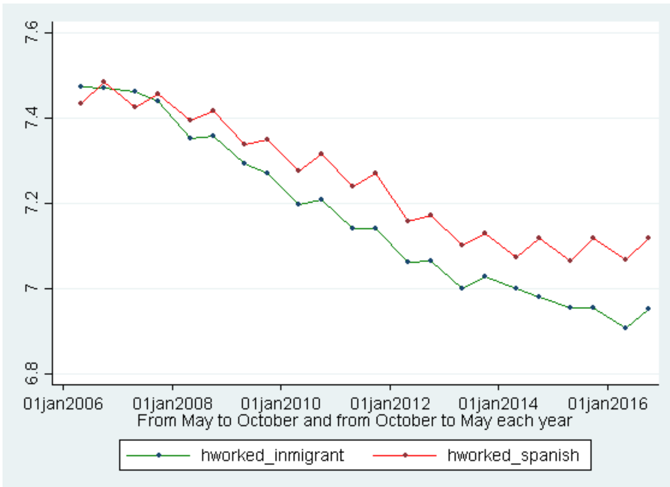
Figure 52: **Monthly Wages within inmigrant groups**



Source: made with Stata.

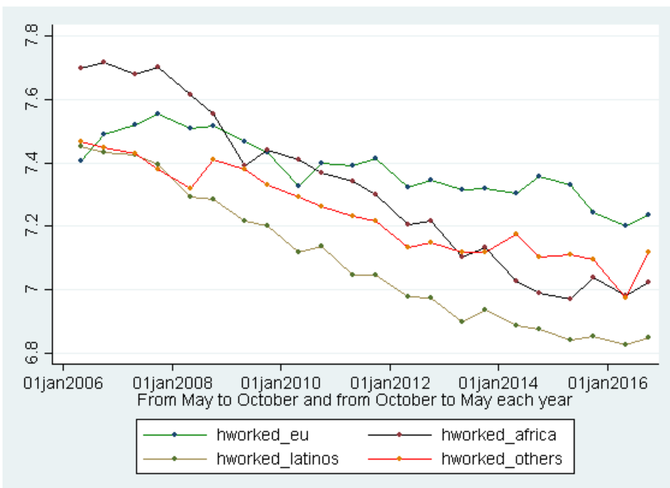
It would also be interesting to display the evolution of worked hours. If we take a look at the trend within natives and immigrants, we can see that is negative, and that the immigrants function is below the native one. I stated earlier that this could be because of the higher percentage of part time/temporal jobs of the immigrants group. By distinguishing within groups, we see that the lowest are, once again, the African and the Latino communities, more proof that the cause maybe the one stated.

Figure 53: **Hourly Wages**



Source: made with Stata.

Figure 54: **Hourly Wages within inmigrant groups**



Source: made with Stata.

2.3 Regressions

The assignment tell us to make the Hourly Wage Change as $W_t - W_{t-1}$ and to then take logs. This would produce irrational numbers if the wage change was negative. That's why I dropped all irrational numbers from after the variable was made. I loose a lot of observations, but there are enough to take conclusions from the model.[3]

First, I will make a regression that has as independent variable the dummy "inmigrant", that takes value one if the individual is an inmigrant, and 0 otherwise.

$$log(W_t - W_{t-1}) = \beta_0 + \beta_1 inmigrant \tag{2}$$

Apart from this regression, I will make another with the variable in levels:

$$log(W_t) = \beta_0 + \beta_1 inmigrant \tag{3}$$

As we can see in the tables, the independent variable is relevant, and has a positive coefficient. This means that, if the individual is an inmigrant, the difference between the wages will be increased by 0.0134%. In this second regression, the variable has a negative effect: if the individual is an inmigrant, her wages will be diminished by a 0.11%.

Table 6: Regressing log(hourly_wage_change)

	(1)
	l_hourly_wage_change
inmigrant	0.0134^{*}
	(1.99)
Constant	1.065^{***}
	(667.50)
Observations	264212
Adjusted R^2	-0.000

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[3]The regressions will be in the annex.

Table 7: Regressing log(hourly_wage)

	(1)
	ln_h_wage
inmigrant	-0.115***
	(-60.36)
Constant	2.205***
	(4628.80)
Observations	1315659
Adjusted R^2	0.003

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Lets take a look at the results if we regress the same equations but with the inmigrant variable in levels. If the variable takes value 1, the inmigrant is an european, 2 if she is African, and 3 if the nationality of the individual is South American. The ”others” level is taken as the benchmark. In the first regression (the independent variable is Hourly Wage Change), the variable is relevant again, with two of the coefficient of the levels being significant. Now the coefficients are negative, this means that if you are an European or an African individual, the difference between wages (t and $t-1$) will diminish by 0.1 and 0.11%, respectively. In the next regression, the one with $Log(wage_t)$ as independent variable, showcases similar results. The variable is once again relevant for two out of three levels, but this time the level which is not relevant is if you are a European inmigrant. If you are an African, or a South American individual, your hourly wage will diminish 0.17 and 0.17% respectively.

Lets introduce a new variable: hours worked. By adding the new variable to the regressions we made we have that:

- In the first regression, both variables are significant, with inmigrant having a higher effect than before. Now, if you are an inmigrant, the difference between your hourly wages (w_t and w_{t-1}) will diminish by 0.028%, and a 0.18% for each increase in the average of worked hours.

- In the second regression both are again relevant. If you were an inmigrant, your hourly wage would descend a 0.11% and a 0.012% for each worked hours.

- In the third regression the new variable is relevant: an increase of one unit of the average worked hours, will mean a decrease of 0.18% in the difference between wages. Now all the levels are significant. Being an European, African or South America will decrease the difference between the wages in 0.08, 0.10 and 0.04%, respectively.

- For the fourth regression, the new variable is relevant: an increase of one unit will mean a decrease of 0.029% of the hourly wage rate. The categorical variable is relevant in all of its levels, being an European, African or South America will increase your hourly wage rate in 0.015%, or decrease in 0.16 and 0.17%, respectively.

Finally, lets add more covariates to our model: gender, age and work experience. The work experience is a categorical variable that takes value 1 if the individual has less than 6 months of experience, 2 if the experience is between 6-12 months, 3 if it's between 13-24, 4 if it's between 25 and 36, 5 if it's between 37 and 48 months, and 6 if the individual has more than 48 months.

For the regression with Hourly Wage Change as independent variable, we can see that all the variables are relevant, except the South American level of the immigrants categorical variable (which makes a lot of sense, as we saw earlier that the South American individuals have all very high hourly wage rates) and a few of the work experience levels. An increase in the average worked hours, being an inmigrant or being a woman, all decrease the difference in the hourly wage rates. The strongest effects are the gender, and the work experience. The goodness of fit has also increased if we compare it to other instances of the model.

For the regression in levels we find that, once again, the variables are relevant, and the goodness of fit is greatly increase from other instances of the same model. The age has a positive coefficient, as well as the work experience, which is the variable that affects more the outcome. The more experience and the older you are, the higher will be your wage rates. We can find that the categorical variable of immigrants are all negative. This means that being an inmigrant generally carries less hourly wages, on average.

Conclusion

We have withdrawn lots of conclusions from the analysis of the MCVL database: the South Americans have a higher hourly wage rate than the rest, the African people are the most vulnerable group, the EU12 immigrants are the ones that are above all regarding wages (sometimes even more than the natives), etc. From the regressions we got that gender and age were the most prominent variables, as well as the work experience, regarding hourly wage rates. If we take a look at the categorical variable of level of immigrants, we can observe that the coefficients vary wildly from the EU12 inmigrant group, to the African/South America immigrants. For the EU12, the effect of being an inmigrant is much more smaller than the rest of the groups. This means that, on average, the EU12 immigrants have been assimilated much better than the rest of groups. Studies like the IZA paper (Rodríguez-Planas & Nollenberg, 2014), attribute this to the fact that the great majority of individuals within the South American/African groups (the most vulnerable ones) get into mid to low skill jobs, most of them depending on wildly volatile sectors like construction, trade, agriculture or farming. These persons coming from poorer countries often have to get into the first job they can, whereas individuals coming from richer countries have, generally, more years of education and thus, they can get into jobs with a higher skill requirement, that usually are most stable and better paid.

References

[1] Gneezy, U., Niederle, M., Rustichini, A. (2003): *”PERFORMANCE IN COMPETITIVE ENVIRONMENTS: GENDER DIFFERENCES”*, retrieved from `https://egela.ehu.eus/`.

[2] Seguridad Social (2020): *”MCVL Database”*, retrieved from `http://www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/EST211`.

[3] Nollenberg, N., Rodríguez-Planas, N. (2014): *”Labor Market Integration of New Immigrants in Spain”*, retrieved from `http://ftp.iza.org/pp93.pdf`.

ANNEX

Regressions

Table 8: Regressing log(hourly_wage)

	(1)
	l_hourly_wage_change
level_inmigrants=1	-0.118***
	(-4.97)
level_inmigrants=2	-0.105***
	(-3.65)
level_inmigrants=3	0.0260
	(1.51)
level_inmigrants=4	0
	(.)
Constant	1.087***
	(73.92)
Observations	14757
Adjusted R^2	0.003

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Regressing log(hourly_wage)

	(1)
	ln_h_wage
level_inmigrants=1	0.0109
	(1.62)
level_inmigrants=2	-0.170***
	(-22.03)
level_inmigrants=3	-0.170***
	(-35.90)
level_inmigrants=4	0
	(.)
Constant	2.204***
	(540.77)
Observations	81638
Adjusted R^2	0.023

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Regressing log(hourly_wage_change)

	(1)
	l_hourly_wage_change
inmigrant	-0.0285***
	(-4.63)
hworked	-0.182***
	(-231.84)
Constant	2.364***
	(408.25)
Observations	264212
Adjusted R^2	0.169

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Regressing log(hourly_wage)

	(1)
	ln_h_wage
inmigrant	-0.117***
	(-61.33)
hworked	-0.0124***
	(-44.93)
Constant	2.294***
	(1120.23)
Observations	1315659
Adjusted R^2	0.004

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Regressing log(hourly_wage_change)

	(1)
	l_hourly_wage_change
level_inmigrants=1	-0.0879***
	(-4.14)
level_inmigrants=2	-0.107***
	(-4.12)
level_inmigrants=3	-0.0467**
	(-3.01)
level_inmigrants=4	0
	(.)
hworked	-0.184***
	(-59.56)
Constant	2.397***
	(93.43)
Observations	14757
Adjusted R^2	0.196

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Regressing log(hourly_wage)

	(1)
	ln_h_wage
level_inmigrants=1	0.0153*
	(2.29)
level_inmigrants=2	-0.168***
	(-21.85)
level_inmigrants=3	-0.177***
	(-37.50)
level_inmigrants=4	0
	(.)
hworked	-0.0290***
	(-28.10)
Constant	2.413***
	(284.62)
Observations	81638
Adjusted R^2	0.032

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Do file

```
********************************************************************************
********************************************************************************
*                                    PART I                                    *
*                                                                              *
********************************************************************************
********************************************************************************



********************************************************************************
* EXERCISE 1
*
* Report (i) % and absolute number of individuals included in Personal but
* not In Afiliados, (ii) % and absolute number of individuals included in
* both files.Can you explain the reason for having disparities in
* the number of observations of each file?

log using log, replace

clear all
set more off
capture log close

cd "C:\Users\pimpeter\Desktop\LABOR_2\DATA"
use "personal16.dta",clear

sort IDENTPERS

merge IDENTPERS using "afiliados16_1.dta"
tab _merge

* The value of 1 means that they belong only to the master file (Personal)
```

Table 14: Regressing log(hourly_wage_change)

	(1)
	l_hourly_wage_change
hworked	-0.191***
	(-60.68)
level_inmigrants=1	-0.0743***
	(-3.49)
level_inmigrants=2	-0.121***
	(-4.71)
level_inmigrants=3	-0.0240
	(-1.54)
level_inmigrants=4	0
	(.)
Gender	-0.135***
	(-10.68)
exp_cat=1	0
	(.)
exp_cat=2	0.114
	(1.33)
exp_cat=3	0.0722
	(0.98)
exp_cat=4	0.0977
	(1.34)
exp_cat=6	0.198**
	(2.92)
ageOct16	-0.00128
	(-0.97)
Constant	2.491***
	(29.53)
Observations	14757
Adjusted R^2	0.204

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Regressing log(hourly_wage)

	(1)
	ln_h_wage
hworked	-0.0471***
	(-46.31)
level_inmigrants=1	-0.0224***
	(-3.48)
level_inmigrants=2	-0.199***
	(-27.01)
level_inmigrants=3	-0.143***
	(-31.39)
level_inmigrants=4	0
	(.)
Gender	-0.134***
	(-36.79)
exp_cat=1	0
	(.)
exp_cat=2	0.1000***
	(7.72)
exp_cat=3	0.146***
	(13.03)
exp_cat=4	0.185***
	(16.09)
exp_cat=6	0.297***
	(29.62)
ageOct16	0.0166***
	(45.09)
Constant	1.855***
	(109.95)
Observations	81638
Adjusted R^2	0.111

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
* 922,925
* And a value of 3 if they are in both of them
* 6,047,437

sort IDENTPERS FALTA FBAJA
br IDENTPERS dbirth SEXO FALTA FBAJA _merge if _merge==1
* Those who are not matched have no data in FALTA FBAJA

* We loose information because we do not have information for those
* individuals that are not present in the afiliacion database.
* How many individuals are in both databases? We can check this:

tab indiv _merge

********************************************************************************
* EXERCISE 2
*
*
*
*


clear all
set more off
capture log close
cd "C:\Users\pimpeter\Desktop\LABOR_2\DATA"
use afiliacion_personal16_2.dta, replace

tab indiv
* We have 285,207 individuals same as n==N


* Labor Market situation
```

```
* 2016
tab situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
* 2006
tab situlab if FALTA<=mdy(10,1,2006) & FBAJA>=mdy(10,1,2006)



* By Gender
tab situlab SEXO if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), co
tab situlab SEXO if FALTA<=mdy(10,1,2006) & FBAJA>=mdy(10,1,2006), co
*
label define sexo 1"Men" 2"Women"
label values SEXO sexo
tab situlab SEXO if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), co

splitvallabels situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
catplot SEXO situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
percent(SEXO) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(`r(relabel)')) ///
ytitle("%", size(small)) ///
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\hist_gender.png", as(png) replace



* By Age Groups
gen y_end_panel=mdy(10,1,2016)
format y_end_panel %td
gen ageOct16=y_end_panel-dbirth
replace ageOct16=int(ageOct16/365)
gen age_group=.
replace age_group=1 if ageOct16<30
```

```
replace age_group=2 if ageOct16>=30 & ageOct16<44
replace age_group=3 if ageOct16>=44

histogram ageOct16

tab situlab age_group if FALTA<=mdy(10,1,2006) & FBAJA>=mdy(10,1,2006),co
tab situlab age_group if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016),co


tab situlab age_group if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), co

splitvallabels situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
catplot age_group situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
percent(age_group) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(`r(relabel)')) ///
ytitle("%", size(small)) ///
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\hist_age.png", as(png) replace



* By Nationalities and by inmigrant status:
gen inmigrants=.
replace inmigrants=1 if PAISNAC!="N00"
replace inmigrants=0 if PAISNAC=="N00"

gen nationality=.
replace nationality=1 if PAISNAC=="N01" | PAISNAC=="N10" | ///
PAISNAC=="N11" | PAISNAC=="N15" | PAISNAC=="N16" | PAISNAC=="N19"
replace nationality=2 if PAISNAC=="N12" | PAISNAC=="N23"
```

```
replace nationality=3 if PAISNAC=="N06" | PAISNAC=="N03" | PAISNAC=="N07" | ///
PAISNAC=="N08" | PAISNAC=="N09" | PAISNAC=="N13" | PAISNAC=="N22" | ///
PAISNAC=="N26" | PAISNAC=="N27"
tab nationality, miss
replace nationality=4 if nationality==. & PAISNAC!="N00"
replace nationality=5 if PAISNAC=="N00"

tab situlab nationality if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016),co
tab situlab nationality if FALTA<=mdy(10,1,2006) & FBAJA>=mdy(10,1,2006),co

tab situlab inmigrants if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016),co
tab situlab inmigrants if FALTA<=mdy(10,1,2006) & FBAJA>=mdy(10,1,2006),co



label define natgroups 1"EU12" 2"Africa" 3"South America" 4"Others" 5 "Spanish", replace
label values nationality natgroups
tab situlab nationality if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), co

splitvallabels situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
catplot nationality situlab if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
percent(nationality) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(‘r(relabel)’)) ///
ytitle("%", size(small)) ///
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\hist_nation.png", as(png) replace
```

```
* Overall
graph bar if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
over(SEXO) over(situlab) blabel(bar)

* by inmigrants
graph bar if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
over(inmigrants) over (situlab) blabel(bar)


save "ex2.dta", replace

********************************************************************************
* EXERCISE 3:
*
clear all
set more off
capture log close
cd "C:\Users\pimpeter\Desktop\LABOR_2\DATA"
use ex2.dta, replace

gen hworked2=(COEFPARC/1000)*8
replace hworked2=8 if COEFPARC==0
drop hworked
rename hworked2 hworked


* Employed individuals only: contract type

* All
tab tipo if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
```

```
* By gender
tab tipo SEXO if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016),co
* AGE
tab tipo age_group if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016),co
* NATIONALITY
tab tipo nationality if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016),co


* Employed individuals only: worked hours

* All
tab tipo if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), sum (hworked)
* By gender
tab tipo SEXO if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), sum (hworked)
* AGE
tab tipo age_group if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), sum (hworked)
* NATIONALITY
tab tipo nationality if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), sum (hworked)
*****************


* BY GENDER
tab tipo SEXO if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
catplot tipo SEXO if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
percent(SEXO) asyvars stack bar(1, bcolor(lime)) bar(2, bcolor(dkgreen)) ///
bar(3, bcolor(sienna)) bar(4, bcolor(brown))
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\h_gender_t.png", as(png) replace


* BY AGE
tab tipo age_group if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
catplot tipo age_group if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
percent(age_group) asyvars stack bar(1, bcolor(lime)) bar(2, bcolor(dkgreen)) ///
```

```
bar(3, bcolor(sienna)) bar(4, bcolor(brown))
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\h_age_t.png", as(png) replace


* BY NATIONALITY
tab tipo nationality if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016)
catplot tipo nationality if FALTA<=mdy(10,1,2016) & FBAJA>=mdy(10,1,2016), ///
percent(nationality) asyvars stack bar(1, bcolor(lime)) bar(2, bcolor(green)) ///
bar(3, bcolor(sienna)) bar(4, bcolor(nrown))
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\h_nation_t.png", as(png) replace




********************************************************************************
clear all
set more off
capture log close


cd "C:\Users\pimpeter\Desktop\LABOR_2\DATA"
use "data.dta",clear

label var age_oct16 "Age of the individual at October 2016"
label var age_cycle "Age of the individual at that cycle"
label var age_cycle "Age of the individual at that cycle"
label var minFALTAwm "Labor experience-Starting working day"

save, replace
* 108,564 individuals


* We have to include data from residence.
```

```
use SDFMCVL2016PERSONAL.dta, clear
keep IDENTPERS DOMICILIO PROVPRIAF

sort IDENTPERS
save provinces_info.dta, replace


use data.dta, clear
sort IDENTPERS
rename _merge merge
merge m:1 IDENTPERS using provinces_info.dta


save data_2.dta, replace

eststo clear
eststo: regress hworked i.age_group
esttab using regols.tex, ar2 label  ///
title("Regressing Worked Hours by Age groups")

eststo clear
eststo: regress hworked i.nationality
esttab using regols2.tex, ar2 label  ///
title("Regressing Worked Hours by Nationalities")

eststo clear
eststo: regress hworked i.SEXO
esttab using regols3.tex, ar2 label  ///
title("Regressing Worked Hours by Gender")
********************************************************************************
* EXER. 4
clear all
set more off
```

```
capture log close


cd "C:\Users\pimpeter\Desktop\LABOR_2\DATA"
use "data_2.dta",clear

label var FALTA "Staring date of afiliation episode on the Social Security"
label var FBAJA "Ending date of afiliation episode on the Social Security"

*Compute total time as employed, self-employed and unemployed at each t
*with using current_empdur current_selfempdur current_unempdur

label var wage "Proxy of wages for each cycle"
label var year "Year"
label var nindef_t "Number of indefinite contracts in each cycle per individual"
label var ntemp_t "Number of temporary contracts in each cycle per individual"
label var time_emp "Number of temporary contracts in each cycle per individual"

save, replace

* I have 108564 individuals after I dropped out 6713 (so, we had 115,277)

drop if indiv==.

* Number of total observations
display 108564*22

des

********************************************************************************
* EXERCISE 5:
*
* MERGING CONVIVIENTES FOR NEXT PART
```

```
clear all
set more off
capture log close


cd "C:\Users\pimpeter\Desktop\LABOR_2\DATA"
use "data_2.dta",clear

drop _merge
sort IDENTPERS
merge IDENTPERS using "convivientes16.dta"

tab _merge

*rename DOMICILIO residence

save, replace

********************************************************************************

bysort IDENTPERS: gen hworked2=8 if hworked==8
bysort IDENTPERS: replace hworked2=8-hworked
bysort IDENTPERS: replace hworked2=8 if hworked2==0
drop hworked
rename hworked2 hworked

save, replace


********************************************************************************

********************************************************************************
********************************************************************************
```

```
*                                  PART II                                   *
*                                                                            *
******************************************************************************
******************************************************************************


******************************************************************************
* EXERCISE 1:
* A)

* Exercise 1a)
tab gender if indiv==1

* Age at the start of the time period:

gen year2006=mdy(1,1,2006)
format year2006 %td
gen agejan06=year2006-dbirth
replace agejan06=int(agejan06/365)
drop year2006
histogram ageOct16 if indiv==1, ///
percent fcolor(mint) barwidth(1) gap(7) addlabel

*RESIDENCE
replace residence=int(residence/1000)
tab residence

gen inmigrant=1 if paisnac!="N00"
replace inmigrant=0 if paisnac=="N00"

* NATIONALITY

gen nationality=.
```

```
replace nationality=1 if paisnac=="N01" | paisnac=="N10" | ///
paisnac=="N11" | paisnac=="N15" | paisnac=="N16" | paisnac=="N19"
replace nationality=2 if paisnac=="N12" | paisnac=="N23"
replace nationality=3 if paisnac=="N06" | paisnac=="N03" | paisnac=="N07" | ///
paisnac=="N08" | paisnac=="N09" | paisnac=="N13" | paisnac=="N22" | ///
paisnac=="N26" | paisnac=="N27"
replace nationality=4 if nationality==. & paisnac!="N00"
replace nationality=5 if paisnac=="N00"

tab nationality if indiv==1
********************************************************************************
* EXERCISE 1:
* B)
* CREATING VARIABLES



* SITLAB: LABOR SITUATION


* TYPE OF CONTRACT: TIPO


* HOURS WORKED:
bysort t: egen avg_hworked=mean(hworked)



* MONTHLY WAGES:
bysort t: egen monthly_wages=mean(wage)
bysort t: egen monthly_wages_eu12=mean(wage) if nationality==1
bysort t: egen monthly_wages_africa=mean(wage) if nationality==2
bysort t: egen monthly_wages_latinos=mean(wage) if nationality==3
bysort t: egen monthly_wages_others=mean(wage) if nationality==4
bysort t: egen monthly_wages_spain=mean(wage) if nationality==5
* HOURLY WAGES:
```

```
bysort IDENTPERS: gen wage_day=(wage/22) if situlab==1
bysort IDENTPERS: gen hourly_wage=(wage_day/hworked) if situlab==1
drop wage_day
bysort t: egen hourly_wages_eu12=mean(hourly_wage) if nationality==1
bysort t: egen hourly_wages_africa=mean(hourly_wage) if nationality==2
bysort t: egen hourly_wages_latinos=mean(hourly_wage) if nationality==3
bysort t: egen hourly_wages_others=mean(hourly_wage) if nationality==4
bysort t: egen hourly_wages_spain=mean(hourly_wage) if nationality==5

*----------

* LABOR SITUATION BY NATIONALITY

tab situlab nationality if t==1 & indiv==1,co
tab situlab nationality if t==22,co

tab situlab age_group if t==1

label define natgroups 1 "EU12" 2"Africa" 3"South America" 4"Others" 5 "Spain", replace
label values nationality natgroups

splitvallabels situlab if t==1
catplot nationality situlab if t==1, ///
percent(nationality) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(‘r(relabel)’)) ///
ytitle("%", size(small)) ///
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\1.png", as(png) replace

splitvallabels situlab if t==22
```

```
catplot nationality situlab if t==22, ///
percent(nationality) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(`r(relabel)')) ///
ytitle("%", size(small)) ///
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\2.png", as(png) replace



* TYPE OF CONTRACT
tab tipo nationality if t==1 & indiv==1,co
tab tipo nationality if  t==22,co



splitvallabels tipo if t==1
catplot nationality tipo if t==1, ///
percent(nationality) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(`r(relabel)')) ///
ytitle("%", size(small)) ///
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\3.png", as(png) replace

splitvallabels tipo if t==22
catplot nationality tipo if t==22, ///
percent(nationality) ///
var1opts(label(labsize(small))) ///
var2opts(label(labsize(small)) relabel(`r(relabel)')) ///
ytitle("%", size(small)) ///
```

```
blabel(bar, format(%4.1f)) ///
intensity(25) ///
asyvars
graph export "C:\Users\pimpeter\Desktop\LABOR_2\figures\4.png", as(png) replace



* HOURS WORKED
tab nationality if t==1, sum (hworked)
tab nationality if t==22, sum (hworked)



* MONTHLY WAGES

sum monthly_wages_eu12 monthly_wages_africa monthly_wages_latinos ///
monthly_wages_others monthly_wages_spain if t==1

sum monthly_wages_eu12 monthly_wages_africa monthly_wages_latinos ///
monthly_wages_others monthly_wages_spain if t==22





* HOURLY WAGES
sum hourly_wages_eu12 hourly_wages_africa hourly_wages_latinos ///
hourly_wages_others hourly_wages_spain if t==1

sum hourly_wages_eu12 hourly_wages_africa hourly_wages_latinos ///
hourly_wages_others hourly_wages_spain if t==22
```

```
* OTHERS
*latinosvsus spanish monthly wages
twoway (line monthly_wages_spain cycle, sort) ///
(line monthly_wages_latinos cycle, sort)

graph set window fontface "LM Roman 10"

*latinos vsus spanish hourly wages
twoway (line hourly_wages_spain cycle, sort) ///
(line hourly_wages_latinos cycle, sort)

graph set window fontface "LM Roman 10"


*regression
eststo clear
eststo: reg hourly_wage ib5.nationality i.tipo hworked
esttab using rege.tex, ar2 label  ///
title("Regressing Hourly Wages")


********************************************************************************
* EXERCISE 2:

* INMIGRANTS

gen inmigrants=.
replace inmigrants=1 if paisnac!="N00"
replace inmigrants=0 if paisnac=="N00"

bysort t: egen hourly_wages_inmigrant=mean(hourly_wage) if inmigrants==1
bysort t: egen hourly_wages_spanish=mean(hourly_wage) if inmigrants==0
```

```
bysort t: egen monthly_wages_inmigrant=mean(monthly_wages) if inmigrants==1
bysort t: egen monthly_wages_spanish=mean(monthly_wages) if inmigrants==0

* hourly graph
twoway (connected hourly_wages_inmigrant cycle, sort lcolor(green) lwidth(medium) lpattern(solid)
(connected hourly_wages_spanish cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(smal

* monthly graph

twoway (connected monthly_wages_eu12 cycle, sort lcolor(green) lwidth(medium) lpattern(solid) msi
(connected monthly_wages_spanish cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(sma
(connected monthly_wages_africa cycle, sort lcolor(black) lwidth(medium) lpattern(solid) msize(sm
(connected monthly_wages_latinos cycle, sort lcolor(brown) lwidth(medium) lpattern(solid) msize(s
(connected monthly_wages_others cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(smal

* 4 TYPES OF INMIGRANTS

* hourly graph
twoway (connected hourly_wages_eu12 cycle, sort lcolor(green) lwidth(medium) lpattern(solid) msiz
(connected hourly_wages_africa cycle, sort lcolor(black) lwidth(medium) lpattern(solid) msize(sma
(connected hourly_wages_latinos cycle, sort lcolor(brown) lwidth(medium) lpattern(solid) msize(sm
(connected hourly_wages_others cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(small


* monthly graph
twoway (connected monthly_wages_eu12 cycle, sort lcolor(green) lwidth(medium) lpattern(solid) msi
(connected monthly_wages_africa cycle, sort lcolor(black) lwidth(medium) lpattern(solid) msize(sm
(connected monthly_wages_latinos cycle, sort lcolor(brown) lwidth(medium) lpattern(solid) msize(s
(connected monthly_wages_others cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(smal


********************************************************************************
* EXERCISE 3:
```

```
* INMIGRANTS
bysort t: egen hworked_inmigrant=mean(hworked) if inmigrants==1
bysort t: egen hworked_spanish=mean(hworked) if inmigrants==0

* hworked graph
twoway (connected hworked_inmigrant cycle, sort lcolor(green) lwidth(medium) lpattern(solid) msiz
(connected hworked_spanish cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(small))

* % unemployment
tab situlab if inmigrants==1
tab situlab if inmigrants==0

* 4 TYPES OF INMIGRANTS
* monthly graph
bysort t: egen hworked_eu=mean(hworked) if nationality==1
bysort t: egen hworked_africa=mean(hworked) if nationality==2
bysort t: egen hworked_latinos=mean(hworked) if nationality==3
bysort t: egen hworked_others=mean(hworked) if nationality==4

twoway (connected hworked_eu cycle, sort lcolor(green) lwidth(medium) lpattern(solid) msize(small
(connected hworked_africa cycle, sort lcolor(black) lwidth(medium) lpattern(solid) msize(small))
(connected hworked_latinos cycle, sort lcolor(brown) lwidth(medium) lpattern(solid) msize(small))
(connected hworked_others cycle, sort lcolor(red) lwidth(medium) lpattern(solid) msize(small))



********************************************************************************
* EXERCISE 4:
********************************************************************************
********************************************************************************
************ REGRESIONS
********************************************************************************
```

```
*******************************************************************************

* A) HOURLY WAGE CHANGE



tabstat hourly_wage, statistics(mean median sd p25 p50 p75)


* Creating LN wage
bysort IDENTPERS: gen ln_h_wage=log(hourly_wage)
bysort IDENTPERS: gen ln_hourly_wage_rate =ln_h_wage[_n]-ln_h_wage[_n-1]
bysort IDENTPERS: gen hourly_wage_change =hourly_wage[_n]-hourly_wage[_n-1]
bysort IDENTPERS: gen l_hourly_wage_change =log(hourly_wage_change)

replace l_hourly_wage_change=. if l_hourly_wage_change<0

*regresiones
eststo clear
eststo: xtreg l_hourly_wage_change inmigrant, i(cycle) fe
esttab using r1.tex, ar2 label  ///
title("Regressing log(hourly_wage_change)")

eststo clear
eststo: xtreg ln_h_wage inmigrant, i(cycle) fe
esttab using r2.tex, ar2 label  ///
title("Regressing log(hourly_wage)")




* B) HOURLY WAGE CHANGE
*
gen level_inmigrants= nationality
replace level_inmigrants=. if nationality==5
```

```
eststo clear
eststo: xtreg ln_h_wage ib4.level_inmigrants, i(cycle) fe
esttab using r3.tex, ar2 label  ///
title("Regressing log(hourly_wage)")


eststo clear
eststo: xtreg l_hourly_wage_change ib4.level_inmigrants, i(cycle) fe
esttab using r4.tex, ar2 label  ///
title("Regressing log(hourly_wage)")








*-----------------------------------


* C) INCLUDING HOURS WORKED IN REGRESSIONS A & B
*
* a
eststo clear
eststo: xtreg ln_h_wage inmigrant hworked, i(cycle) fe
esttab using r5.tex, ar2 label  ///
title("Regressing log(hourly_wage)")

eststo clear
eststo: xtreg l_hourly_wage_change inmigrant hworked, i(cycle) fe
esttab using r6.tex, ar2 label  ///
```

```
title("Regressing log(hourly_wage_change)")

* b

eststo clear
eststo: xtreg ln_h_wage ib4.level_inmigrants hworked, i(cycle) fe
esttab using r7.tex, ar2 label  ///
title("Regressing log(hourly_wage)")

eststo clear
eststo: xtreg l_hourly_wage_change ib4.level_inmigrants hworked, i(cycle) fe
esttab using r8.tex, ar2 label  ///
title("Regressing log(hourly_wage_change)")




* D) LABOUR EXPERIENCE CATEGORIES
* Removing negative experience
*Negative values in experience
bysort IDENTPERS: gen experience=labor_experience
forval x=1/22{
set more off
bysort IDENTPERS: replace experience=experience[_n-`x'] if experience<0
}


* Generating variable with experience categories
```

```
gen exp_cat =.
replace exp_cat = 1 if labor_experience <= 182
replace exp_cat = 2 if labor_experience > 182 & labor_experience <= 365
replace exp_cat = 3 if labor_experience > 365 & labor_experience <= 730
replace exp_cat = 4 if labor_experience > 730 & labor_experience <= 1095
replace exp_cat = 5 if labor_experience > 1095 & labor_experience <= 1040
replace exp_cat = 6 if labor_experience > 1040
* Regression
gen y_end_panel=mdy(10,1,2016)
format y_end_panel %td
gen ageOct16=y_end_panel-dbirth
replace ageOct16=int(ageOct16/365)

eststo clear
eststo: xtreg ln_h_wage hworked ib4.level_inmigrants gender i.exp_cat ageOct16, i(cycle) fe
esttab using r9.tex, ar2 label  ///
title("Regressing log(hourly_wage)")

eststo clear
eststo: xtreg l_hourly_wage_change hworked ib4.level_inmigrants gender i.exp_cat ageOct16, i(cycl
esttab using r10.tex, ar2 label  ///
title("Regressing log(hourly_wage_change)")


xtreg ln_h_wage hworked ib4.level_inmigrants gender i.exp_cat ageOct16, i(cycle) fe
xtreg ln_hourly_wage_rate hworked ib4.level_inmigrants gender i.exp_cat ageOct16, i(cycle) fe
log close
```