

Tarea 2 Panel SanMartín_Cortes

May 23, 2023

0.0.1 Housekeeping and Data

1 Section 3: Panel data

```
[159]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
import sklearn
import scipy
import linearmodels.panel as lmp
import pytwoway as tw
import bipartitepandas as bpd
import seaborn as sns
from linearmodels.iv import IV2SLS

%matplotlib inline
```

```
[161]: enia.export = enia.export.astype(int)
```

Las variables tienen la siguiente descripción:

- *ID*: firm unique identifier
- *year*: survey year
- *tamano*: 1 large, 2 medium, 3 small, 4 micro (funcion de las ventas y el numero de trabajadores)
- *sales*: sales (in log of 1,000 CLP)
- *age*: firm age at time of survey
- *foreign*: non-domestic firm (binary)
- *export*: production for export (binary)
- *workers*: log of number of workers

- *fomento*: firm receives public incentives (binary)
- *iyd*: firm does I+D (binary)
- *impuestos*: taxes (in million US)
- *utilidades*: firm revenue (in million US)

[162]: `enia.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39104 entries, 0 to 39103
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   ID              39104 non-null  int64
1   year            39104 non-null  int64
2   tamano          39104 non-null  int64
3   sales           39104 non-null  float64
4   age             39104 non-null  int64
5   foreign         39104 non-null  int64
6   export          39104 non-null  int64
7   workers         39104 non-null  float64
8   fomento         39104 non-null  int64
9   iyd             39104 non-null  int64
10  impuestos       39104 non-null  float64
11  utilidades      39104 non-null  float64
dtypes: float64(4), int64(8)
memory usage: 3.6 MB
```

1.1 Pregunta 1

Cargar la base de datos *enia.csv* en el ambiente. Identifique los tipos de datos que se encuentran en la base, realice estadísticas descriptivas sobre las variables importantes (Hint: Revisar la distribución, datos faltantes, outliers, etc.) y limpie las variables cuando sea necesario. Para las preguntas 2-8 **EXCLUYA LA VARIABLE FOMENTO DE SU ANALISIS**.

[163]: `enia.describe()`

```
[163]:
```

	ID	year	tamano	sales	age \
count	39104.000000	39104.000000	39104.000000	39104.000000	39104.000000
mean	218089.455554	2011.787183	2.248773	3.574172	15.305084
std	128228.474792	3.781237	1.153089	1.692742	12.488330
min	100000.000000	2007.000000	1.000000	0.000000	0.000000
25%	105409.000000	2007.000000	1.000000	2.337643	7.000000
50%	200994.500000	2013.000000	2.000000	3.553321	14.000000
75%	302466.250000	2015.000000	3.000000	4.539098	20.000000
max	507526.000000	2017.000000	4.000000	10.309005	190.000000

	foreign	export	workers	fomento	iyd \
count	39104.000000	39104.000000	39104.000000	39104.000000	39104.000000
mean	0.081859	0.111191	1.757726	0.076105	0.224887
std	0.274153	0.314372	1.186507	0.265169	0.417514
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.778151	0.000000	0.000000
50%	0.000000	0.000000	1.785330	0.000000	0.000000
75%	0.000000	0.000000	2.661813	0.000000	0.000000
max	1.000000	1.000000	5.845915	1.000000	1.000000

	impuestos	utilidades
count	39104.000000	3.910400e+04
mean	0.203856	1.875255e+00
std	15.869466	2.306899e+02
min	-180.992528	-2.443698e+02
25%	0.000000	9.050000e-07
50%	0.000007	8.080000e-05
75%	0.000167	1.283704e-03
max	2981.494528	4.544069e+04

Dentro de las variables que se tienen en la tabla de datos están las variables dicotómicas, las cuales son foreign, export, fomento, IyD.

En las variables categoricas está, ID, year.

Por último en la variable numérica están, tamaño, sales, age, workers, impuestos y utilidades.

Luego de ver la información (Sección de Anexos) se estima que la información proporcionada tiene datos relevante por lo que no se eliminará información.

```
[164]: enia = pd.read_csv('../data/enia.csv')
enia.dropna(inplace=True)
enia.reset_index(drop=True, inplace=True)

enia.head()
#variable construction
enia['utilidades']=enia['utilidades']/1000 #acá se divide porque es muy grande
    ↪ el valor
X=enia[['tamano','sales','age','foreign','export','fomento','iyd','impuestos','utilidades']]
Xm=enia.drop('workers', axis=1).groupby(by='ID').transform('mean').drop('year',
    ↪ axis=1)
Xm.columns = ['mtamano','msales', 'mage', 'mforeign', 'mexport', 'mfomento',
    ↪ 'miyd', 'mimpuestos', 'mutilidades' ]
Xc=pd.DataFrame(np.c_[enia, Xm], columns=enia.columns.tolist() + Xm.columns.
    ↪ tolist())
Xc['year'] = Xc['year'].astype(int)
Xc['ID'] = Xc['ID'].astype(int)
#set panel structure
```

```
Xc = Xc.set_index(["ID", "year"])
Xc
```

```
[164]:
```

		tamano	sales	age	foreign	export	workers	fomento	iyd	\
ID	year									
100003	2007	1.0	7.046558	22.0	1.0	1.0	3.486855	0.0	1.0	
	2009	1.0	7.875563	24.0	1.0	1.0	3.504607	0.0	0.0	
	2013	1.0	7.437399	23.0	1.0	1.0	4.691621	1.0	1.0	
	2015	1.0	7.356472	30.0	1.0	1.0	4.682614	1.0	0.0	
	2017	1.0	4.014772	32.0	1.0	1.0	4.611691	1.0	0.0	
...		
507481	2017	4.0	1.401027	17.0	0.0	0.0	0.000000	0.0	0.0	
507482	2017	4.0	1.528097	22.0	0.0	0.0	0.000000	0.0	0.0	
507483	2017	4.0	1.679748	21.0	0.0	0.0	1.397940	0.0	0.0	
507484	2017	4.0	1.285538	9.0	0.0	0.0	1.568202	0.0	0.0	
507487	2017	4.0	1.441296	12.0	0.0	0.0	0.000000	0.0	0.0	
ID	year	impuestos	utilidades	mtamano	msales	mage	mforeign	\		
100003	2007	1.231345	7.113892e-03	1.0	6.746153	26.2	1.0			
	2009	8.762233	3.397611e-02	1.0	6.746153	26.2	1.0			
	2013	0.001886	3.137265e-03	1.0	6.746153	26.2	1.0			
	2015	0.411670	1.298413e-03	1.0	6.746153	26.2	1.0			
	2017	0.000721	1.949247e-06	1.0	6.746153	26.2	1.0			
...				
507481	2017	0.000001	2.240000e-09	4.0	1.401027	17.0	0.0			
507482	2017	0.000000	1.680000e-08	4.0	1.528097	22.0	0.0			
507483	2017	0.000003	1.420000e-09	4.0	1.679748	21.0	0.0			
507484	2017	0.000000	3.520000e-09	4.0	1.285538	9.0	0.0			
507487	2017	0.000002	-3.960000e-09	4.0	1.441296	12.0	0.0			
ID	year	mexport	mfomento	miyd	mimpuestos	mutilidades				
100003	2007	1.0	0.6	0.4	2.081571	9.105525e-03				
	2009	1.0	0.6	0.4	2.081571	9.105525e-03				
	2013	1.0	0.6	0.4	2.081571	9.105525e-03				
	2015	1.0	0.6	0.4	2.081571	9.105525e-03				
	2017	1.0	0.6	0.4	2.081571	9.105525e-03				
...					
507481	2017	0.0	0.0	0.0	0.000001	2.240000e-09				
507482	2017	0.0	0.0	0.0	0.000000	1.680000e-08				
507483	2017	0.0	0.0	0.0	0.000003	1.420000e-09				
507484	2017	0.0	0.0	0.0	0.000000	3.520000e-09				
507487	2017	0.0	0.0	0.0	0.000002	-3.960000e-09				

[39104 rows x 19 columns]

1.2 Pooled OLS

1.2.1 Pregunta 2.

Ejecute un modelo Pooled OLS para explicar el numero de trabajadores. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

Interpretación A contiación se observa la tabla de Pooled OLS. La variables elegidas se basaron en el anexo N°12. En base a esto se observa que el modelo escogido explica el 51%. - Las variables escogidas son significativas para la estimación. - Mientras más pequeña la empresa, más ventas y más utilidades hay existe una menor cantidad de trabajadores, 70%, 3% y 5% respectivamente lo que indica que mientras la empresa es más pequeña necesita menos trabajadores, y si las variables economicas aumentan probablemente automatizan procesos por lo que tienen una reducción de personal. - Por otro lado, la edad, la exportación, si la empresa es extranjera y la aplicación de IyD en la empresa aumenta la cantidad de trabajadores en un 1.3%, 7%, 8% y 11%. lo que se puede intepretar como que si presenta características internacionales mantienen y aumentan la cantidad. trabajadores.

```
[180]: y=Xc['workers'];
X=Xc[['tamano', 'sales', "age", 'utilidades', 'export', 'foreign', 'iyd']];
X=sm.add_constant(X);
model=lm.PooledOLS(y,X);
Res_OLS=model.fit(cov_type="robust");
print(Res_OLS);
```


PooledOLS Estimation Summary

```
=====
Dep. Variable:          workers    R-squared:                0.5104
Estimator:              PooledOLS  R-squared (Between):      0.5496
No. Observations:       39104      R-squared (Within):       0.0830
Date:                   Thu, May 11 2023  R-squared (Overall):      0.5104
Time:                   18:07:18    Log-likelihood            -4.821e+04
Cov. Estimator:         Robust

                               F-statistic:          5821.5
Entities:                24128    P-value                  0.0000
Avg Obs:                  1.6207  Distribution:             F(7,39096)
Min Obs:                  1.0000
Max Obs:                  5.0000  F-statistic (robust):     6436.7
                               P-value                  0.0000
Time periods:              5      Distribution:             F(7,39096)
Avg Obs:                   7820.8
Min Obs:                   6480.0
Max Obs:                   1.021e+04
```

Parameter Estimates

```
=====
Parameter  Std. Err.    T-stat    P-value    Lower CI    Upper CI
-----
const      3.2362      0.0250    129.39     0.0000     3.1872     3.2853
```

tama 	-0.7056	0.0052	-135.48	0.0000	-0.7158	-0.6954
sales	-0.0358	0.0033	-10.991	0.0000	-0.0421	-0.0294
age	0.0129	0.0004	32.014	0.0000	0.0121	0.0136
utilidades	-0.0505	0.0034	-15.016	0.0000	-0.0571	-0.0439
export	0.0709	0.0149	4.7614	0.0000	0.0417	0.1001
foreign	0.0808	0.0185	4.3781	0.0000	0.0446	0.1170
iyd	0.1104	0.0102	10.852	0.0000	0.0905	0.1303

=====

1.3 First differences

El modelo tiene una capacidad de ajuste relativamente baja, explica solo un 19.61%. Es posible que se necesiten considerar otros factores para obtener una explicación más completa en reacción a los trabajadores. - La variable exportación e impuestos son no significativas. - Las ventas disminuyen la cantidad de trabajadores en un 15,28%. - En relación a la edad y utilidades este muestra un aumento de 0,4% y 2,96%. Estos valores son debidos a los efectos que no son fijos.

```
[187]: y=Xc['workers']
X=Xc[['sales', 'age', 'utilidades', 'export', 'impuestos']]
model=lmf.FirstDifferenceOLS(y,X)
fd=model.fit(cov_type="robust")
print(fd)
```

FirstDifferenceOLS Estimation Summary

```
=====
Dep. Variable:          workers    R-squared:          0.1670
Estimator:      FirstDifferenceOLS  R-squared (Between): -0.5233
No. Observations:      14188      R-squared (Within):  0.1475
Date:                Thu, May 11 2023  R-squared (Overall): -0.4923
Time:                18:48:36      Log-likelihood      -1.476e+04
Cov. Estimator:      Robust

                               F-statistic:          568.70
Entities:                24128      P-value           0.0000
Avg Obs:                 1.6207      Distribution:      F(5,14183)
Min Obs:                 1.0000
Max Obs:                 5.0000      F-statistic (robust): 336.58
                               P-value           0.0000
Time periods:                5      Distribution:      F(5,14183)
Avg Obs:                 7820.8
Min Obs:                 6480.0
Max Obs:                 1.021e+04
```

Parameter Estimates

```
=====
Parameter  Std. Err.    T-stat    P-value    Lower CI    Upper CI
-----
sales      -0.1528    0.0038   -39.770    0.0000   -0.1603   -0.1453
age         0.0040    0.0006    6.6338    0.0000    0.0028    0.0052
```

utilidades	0.0296	0.0037	8.0926	0.0000	0.0224	0.0367
export	0.0130	0.0215	0.6017	0.5474	-0.0293	0.0552
impuestos	0.0002	0.0001	1.7542	0.0794	-2.377e-05	0.0004
=====						

1.4 Fixed Effects

1.4.1 Pregunta 3

Ejecute un modelo de efectos fijos para explicar el numero de trabajadores. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado. - El modelo explica el 16,25%, muy cercano al modelo anterior. - Las variables exportación e impuestos son no significativas para el modelo. - Las ventas disminuyen en un 16,26% la cantidad de trabajadores, probablemente se base por las razones mencionadas en el modelo anterior. - Las variables que producen un aumento en la cantidad de trabajadores son edad, utilidades en un 1,2%, 4,3%.

```
[189]: X=Xc[['sales', 'age', 'utilidades', 'export', 'impuestos']]
X=sm.add_constant(X)
model=lm.PanelOLS(y,X, entity_effects=True)
fe=model.fit(cov_type="robust")
print(fe)
```

PanelOLS Estimation Summary

Dep. Variable:	workers	R-squared:	0.1625
Estimator:	PanelOLS	R-squared (Between):	-0.2820
No. Observations:	39104	R-squared (Within):	0.1625
Date:	Thu, May 11 2023	R-squared (Overall):	-0.1727
Time:	18:53:48	Log-likelihood	-1.588e+04
Cov. Estimator:	Robust	F-statistic:	581.12
Entities:	24128	P-value	0.0000
Avg Obs:	1.6207	Distribution:	F(5,14971)
Min Obs:	1.0000	F-statistic (robust):	222.97
Max Obs:	5.0000	P-value	0.0000
Time periods:	5	Distribution:	F(5,14971)
Avg Obs:	7820.8		
Min Obs:	6480.0		
Max Obs:	1.021e+04		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	2.1327	0.0354	60.278	0.0000	2.0634	2.2021
sales	-0.1626	0.0059	-27.385	0.0000	-0.1742	-0.1509
age	0.0131	0.0014	9.1828	0.0000	0.0103	0.0159
utilidades	0.0427	0.0027	15.548	0.0000	0.0373	0.0480

export	0.0548	0.0339	1.6167	0.1060	-0.0116	0.1213
impuestos	-0.0001	0.0002	-0.5108	0.6095	-0.0005	0.0003

=====

F-test for Poolability: 4.4253

P-value: 0.0000

Distribution: F(24127,14971)

Included effects: Entity

1.5 Random Effects

1.5.1 Pregunta 4

Ejecute un modelo de efectos aleatorios para explicar el numero de trabajadores. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado.

- El modelo explica aproximadamente el 38.75% de la variabilidad.
- Al analizar los coeficientes de las variables, se observa que “tamano”, “sales”, tienen coeficientes significativos y negativos. Esto sugiere que a medida que aumentan estos valores, se espera una disminución en el número de trabajadores.
- Por otro lado, las variables “age” muestra un aumento en la cantidad de trabajadores.
- Las variables “utilidades” e “IyD” no muestra una relación significativa con los trabajadores.

```
[190]: model=lmpr.RandomEffects(y,X)
re=model.fit(cov_type="robust")
print(re)
```

RandomEffects Estimation Summary

```
=====
Dep. Variable:          workers    R-squared:                0.0126
Estimator:             RandomEffects  R-squared (Between):      0.1874
No. Observations:      39104    R-squared (Within):       -0.0938
Date:                  Thu, May 11 2023  R-squared (Overall):      0.1647
Time:                  19:11:52    Log-likelihood            -3.879e+04
Cov. Estimator:        Robust

                               F-statistic:          100.17
Entities:              24128    P-value                0.0000
Avg Obs:               1.6207    Distribution:           F(5,39098)
Min Obs:               1.0000
Max Obs:               5.0000    F-statistic (robust):    586.24
                               P-value                0.0000
Time periods:          5    Distribution:           F(5,39098)
Avg Obs:               7820.8
Min Obs:               6480.0
Max Obs:               1.021e+04
```

Parameter Estimates

=====

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	0.9481	0.0142	66.628	0.0000	0.9202	0.9760
sales	0.0795	0.0029	27.019	0.0000	0.0737	0.0852
age	0.0219	0.0006	33.777	0.0000	0.0206	0.0231
utilidades	0.0028	0.0169	0.1675	0.8670	-0.0302	0.0359
export	0.4719	0.0178	26.517	0.0000	0.4370	0.5068
impuestos	-0.0005	0.0005	-1.0273	0.3043	-0.0014	0.0004

```
[191]: re.variance_decomposition
```

```
[191]: Effects                0.600533
      Residual                0.344535
      Percent due to Effects  0.635439
      Name: Variance Decomposition, dtype: float64
```

1.6 Model comparison

1.6.1 Pregunta 5

Comente los resultados obtenidos en 2, 3 y 4. ¿Cuáles y por qué existen las diferencias entre los resultados?. En su opinión, ¿Cuál sería el más adecuado para responder la pregunta de investigación y por qué? ¿Qué variables resultaron ser robustas a la especificación?

- En este caso, el modelo PooledOLS explica alrededor del 51% de la variabilidad de “workers”. El modelo RandomEffects tiene el segundo valor más alto (0.3875), mientras que el modelo Fixed Effects tiene el valor más bajo (0.2217).
- En conclusión, el modelo PooledOLS tiene el mejor ajuste en términos de R-cuadrado y F-statistic, lo que muestra ser el modelo más adecuado para explicar la variabilidad de “workers” en función de las variables utilizadas.
- Todos estos valores posiblemente difieran debido a que en el caso de Fixed Effects se consideran menos variables al igual que en Random Effects, esto debido a que le primero utiliza solo variables que varían en el tiempo, dejando afuera una variable importante como el tamaño de la empresa.
- Se puede concluir que en los 3 el tamaño afecta negativamente la cantidad de trabajadores. La edad y las exportaciones afectan de manera positiva mientras que las otras variables son distintas según el método utilizado.
- En el “Hausman Test” se comparan Fixed Effects y RandomEffects, estos tienen un valor p de 0.0 lo que indica que se puede rechazar la hipótesis nula.

```
[171]: print(lmp.compare({"FE": fe, "RE": re, "Pooled": Res_OLS}))
```

Model Comparison			
	FE	RE	Pooled
Dep. Variable	workers	workers	workers
Estimator	PanelOLS	RandomEffects	PooledOLS
No. Observations	39104	39104	39104

Cov. Est.	Robust	Robust	Robust
R-squared	0.2217	0.3875	0.5100
R-Squared (Within)	0.2217	0.1412	0.0809
R-Squared (Between)	0.2587	0.5419	0.5495
R-Squared (Overall)	0.2788	0.5031	0.5100
F-statistic	710.88	4123.2	6783.5
P-value (F-stat)	0.0000	0.0000	0.0000
=====	=====	=====	=====
const	3.1798	3.4657	3.2480
	(50.870)	(137.31)	(130.88)
tamano	-0.4128	-0.7374	-0.7091
	(-18.128)	(-135.46)	(-137.88)
sales	-0.1795	-0.0874	-0.0352
	(-31.896)	(-30.262)	(-10.814)
age	0.0121	0.0130	0.0128
	(8.9048)	(28.174)	(31.945)
utilidades	0.0393	-0.0017	-0.0505
	(7.5179)	(-0.0824)	(-13.999)
export	0.0398	0.1116	0.0854
	(1.2140)	(7.4431)	(5.8686)
iyd	-0.1846	0.0035	0.1105
	(-10.589)	(0.3656)	(10.854)
=====	=====	=====	=====
Effects	Entity		
-----	-----		

T-stats reported in parentheses

```
[192]: import numpy.linalg as la
        from scipy import stats

        def hausman(fe, re):
            diff = fe.params-re.params
            psi = fe.cov - re.cov
            dof = diff.size -1
            W = diff.dot(la.inv(psi)).dot(diff)
            pval = stats.chi2.sf(W, dof)
            return W, dof, pval

[193]: htest = hausman(fe, re)
        print("Hausman Test: chi-2 = {0}, df = {1}, p-value = {2}".format(htest[0],
        ↪ htest[1], htest[2]))
```

Hausman Test: chi-2 = 2230.203180944794, df = 5, p-value = 0.0

1.7 Correlated Random Effects

1.7.1 Pregunta 6

Ejecute un modelo de efectos aleatorios correlacionados (CRE) para explicar el numero de trabajadores. Seleccione las variables dependientes a incluir en el modelo final e interprete su significado. Es este modelo adecuado, dada la data disponible, para modelar el componente no observado?

- Se puede ver que este modelo explica un 38,75% de las variables del modelo, pero el valor global explica el 50,31%.
- Las variables “age” y “export” son positivas y significativas.
- Las variables “utilidades” e “IyD” son no soificativas.
- Las variables “tamano” y “sales” son negativas.

Si es adecuado, ya que explica un gran porcentaje del modelo, mucho mejor que FE y RE, además su F-statistic es mucho mayor.

```
[197]: X=Xc[['tamano', 'sales', 'age', 'utilidades', 'export', 'iyd']]
X=sm.add_constant(X)
model=lmpr.RandomEffects(y,X)
cre=model.fit(cov_type="robust")
print(cre)
```

RandomEffects Estimation Summary

```
=====
Dep. Variable:          workers    R-squared:                0.3875
Estimator:              RandomEffects    R-squared (Between):    0.5419
No. Observations:      39104    R-squared (Within):      0.1412
Date:                  Thu, May 11 2023    R-squared (Overall):    0.5031
Time:                  19:57:27    Log-likelihood          -3.468e+04
Cov. Estimator:        Robust

                               F-statistic:                4123.2
Entities:              24128    P-value                  0.0000
Avg Obs:               1.6207    Distribution:           F(6,39097)
Min Obs:               1.0000
Max Obs:               5.0000    F-statistic (robust):    5257.6
                               P-value                      0.0000
Time periods:          5    Distribution:           F(6,39097)
Avg Obs:               7820.8
Min Obs:               6480.0
Max Obs:               1.021e+04
```

Parameter Estimates

```
=====
Parameter  Std. Err.    T-stat    P-value    Lower CI    Upper CI
-----
const      3.4657    0.0252    137.31    0.0000    3.4163    3.5152
tamano     -0.7374    0.0054   -135.46    0.0000   -0.7481   -0.7268
sales      -0.0874    0.0029   -30.262    0.0000   -0.0931   -0.0818
age         0.0130    0.0005    28.174    0.0000    0.0121    0.0139
```

utilidades	-0.0017	0.0206	-0.0824	0.9343	-0.0421	0.0387
export	0.1116	0.0150	7.4431	0.0000	0.0822	0.1410
iyd	0.0035	0.0095	0.3656	0.7147	-0.0152	0.0221

=====

```
[198]: print(lmp.compare({"FE": fe, "RE": re, "CRE": cre}))
```

Model Comparison			
	FE	RE	CRE
Dep. Variable	workers	workers	workers
Estimator	PanelOLS	RandomEffects	RandomEffects
No. Observations	39104	39104	39104
Cov. Est.	Robust	Robust	Robust
R-squared	0.1625	0.0126	0.3875
R-Squared (Within)	0.1625	-0.0938	0.1412
R-Squared (Between)	-0.2820	0.1874	0.5419
R-Squared (Overall)	-0.1727	0.1647	0.5031
F-statistic	581.12	100.17	4123.2
P-value (F-stat)	0.0000	0.0000	0.0000
=====	=====	=====	=====
const	2.1327 (60.278)	0.9481 (66.628)	3.4657 (137.31)
sales	-0.1626 (-27.385)	0.0795 (27.019)	-0.0874 (-30.262)
age	0.0131 (9.1828)	0.0219 (33.777)	0.0130 (28.174)
utilidades	0.0427 (15.548)	0.0028 (0.1675)	-0.0017 (-0.0824)
export	0.0548 (1.6167)	0.4719 (26.517)	0.1116 (7.4431)
impuestos	-0.0001 (-0.5108)	-0.0005 (-1.0273)	
tamano			-0.7374 (-135.46)
iyd			0.0035 (0.3656)
=====	=====	=====	=====
Effects	Entity		

T-stats reported in parentheses

1.7.2 Pregunta 7

Usando el modelo CRE, prediga la distribucion del componente no observado. Que puede inferir respecto de la heterogeneidad fija en el tiempo y su impacto en el numero de trabajadores?

1.7.3 Pregunta8

Usando sus respuestas anteriores, que modelo prefiere? que se puede inferir en general respecto del efecto de las variables explicativas sobre el numero de trabajadores?

1.7.4 Pregunta 9

Considere que la variable *fomento* es una politica publica donde aleatoriamente se selecciono un grupo de empresas para recibir recursos financieros dedicados a incentivar I+D. Utilizando fomento como instrumento, estime un modelo en dos etapas para entender el impacto causal de la inversion en I+D sobre el numero de trabajadores, y compare versus el modelo MCO (puntos adicionales para hacerlo en un contexto de panel).

Tarea 2

Instrucciones

Los resultados de los ejercicios propuestos se deben entregar como un notebook por correo electronico a juancaros@udec.cl el dia 9/5 hasta las 21:00.

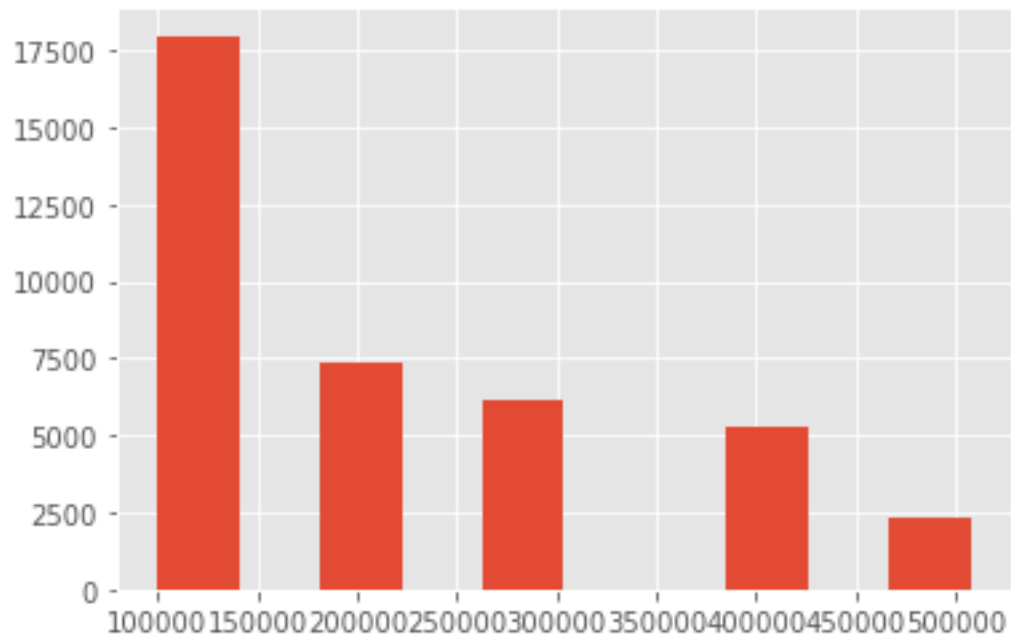
Es importante considerar que el código debe poder ejecutarse en cualquier computadora con la data original del repositorio. Recordar la convencion para el nombre de archivo ademas de incluir en su documento titulos y encabezados por seccion. La data a utilizar es **enia.csv**.

Para este analisis consideraremos tamaño como una variable continua, que identifica el tamaño de la empresa.

1.7.5 Anexo 1

```
[30]: A="ID"
plt.style.use('ggplot')
enia['ID'].hist();
enia.value_counts(A)
```

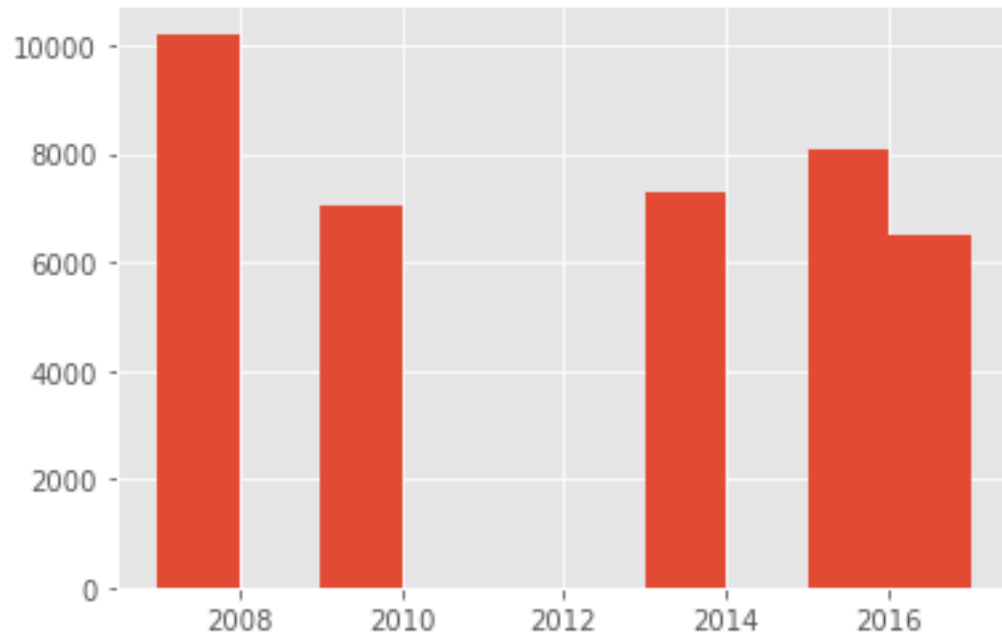
```
[30]: ID
105118    5
103915    5
101364    5
103905    5
105097    5
..
109614    1
109613    1
109612    1
109611    1
507526    1
Length: 24128, dtype: int64
```



1.7.6 Anexo 2

```
[29]: A="year"
plt.style.use('ggplot')
enia['year'].hist();
enia.value_counts(A)
```

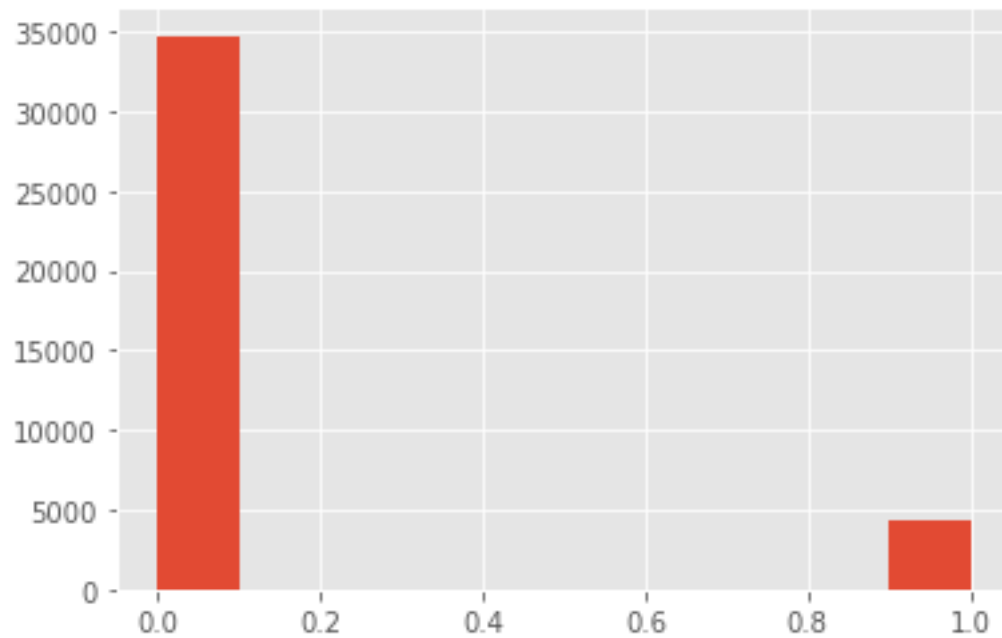
```
[29]: year
2007    10211
2015     8084
2013     7267
2009     7062
2017     6480
dtype: int64
```



1.7.7 Anexo 3

```
[17]: A="export"  
plt.style.use('ggplot')  
enia['export'].hist();  
enia.value_counts(A)
```

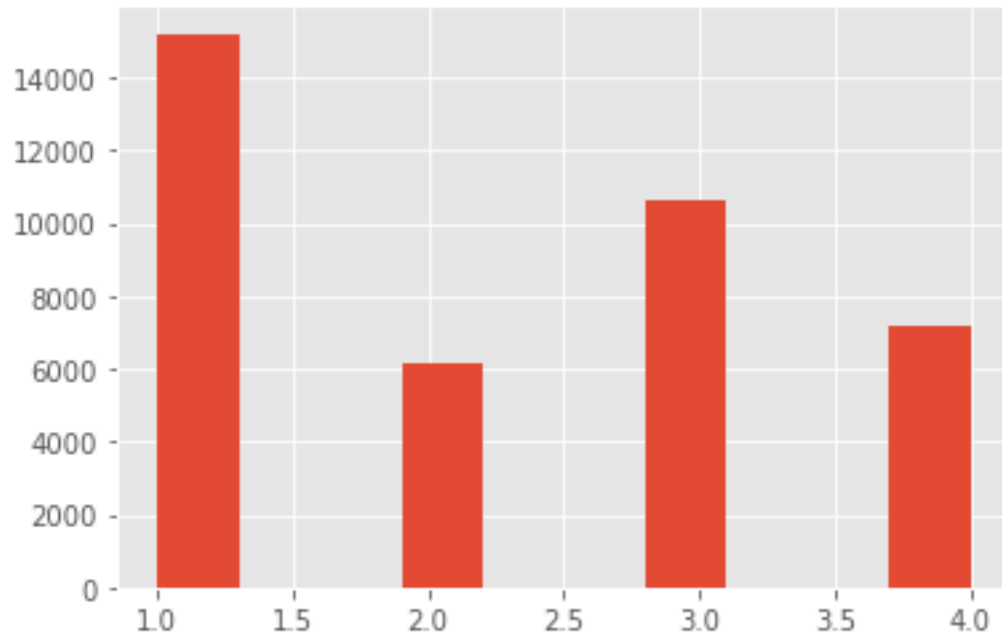
```
[17]: export  
0.0    34756  
1.0     4348  
dtype: int64
```



1.7.8 Anexo 4

```
[18]: A="tamano"
plt.style.use('ggplot')
enia['tamano'].hist();
enia.value_counts(A)
```

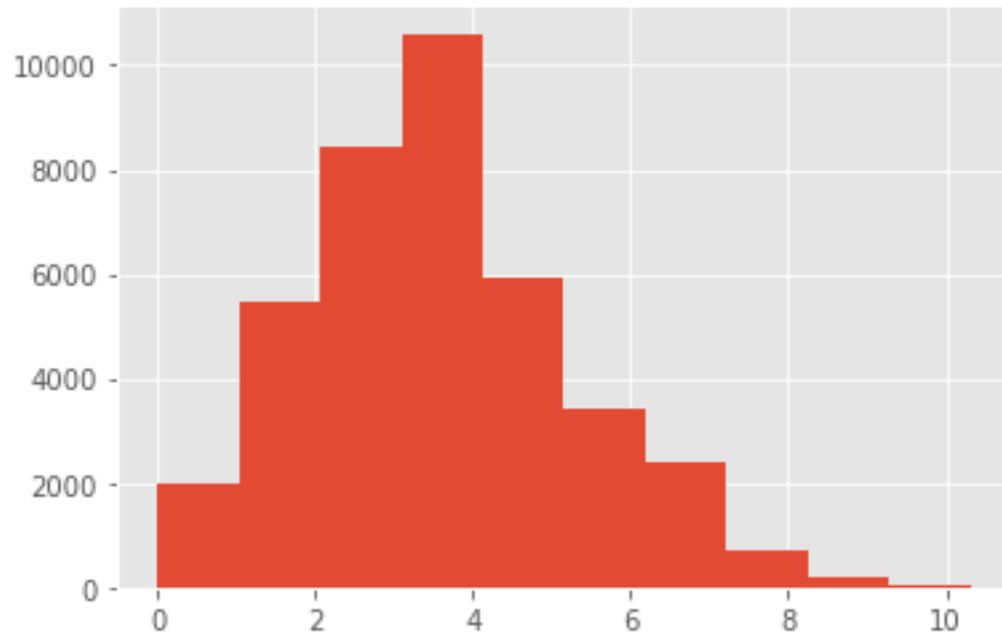
```
[18]: tamano
1    15194
3    10626
4     7148
2     6136
dtype: int64
```

1.7.9 Anexo 5

```
[19]: A="sales"
plt.style.use('ggplot')
enia['sales'].hist();
enia.value_counts(A)
```

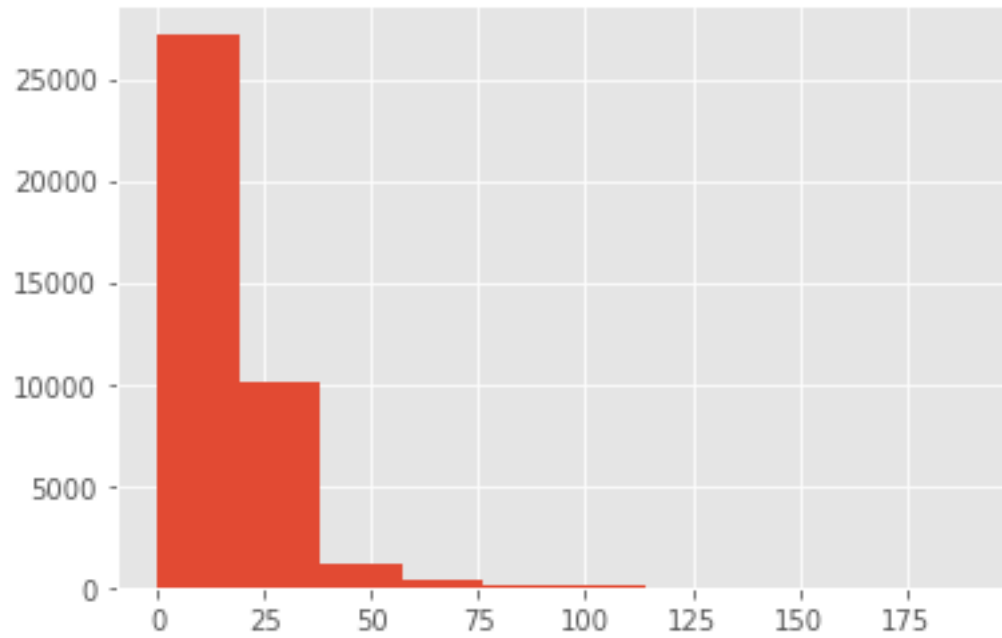
```
[19]: sales
0.000000    652
0.883180     17
1.440375     12
1.154821     12
0.563079     10
...
2.872033      1
2.872068      1
2.872159      1
2.872238      1
10.309005      1
Length: 37781, dtype: int64
```



1.7.10 Anexo 6

```
[20]: A="age"
plt.style.use('ggplot')
enia['age'].hist();
enia.value_counts(A)
```

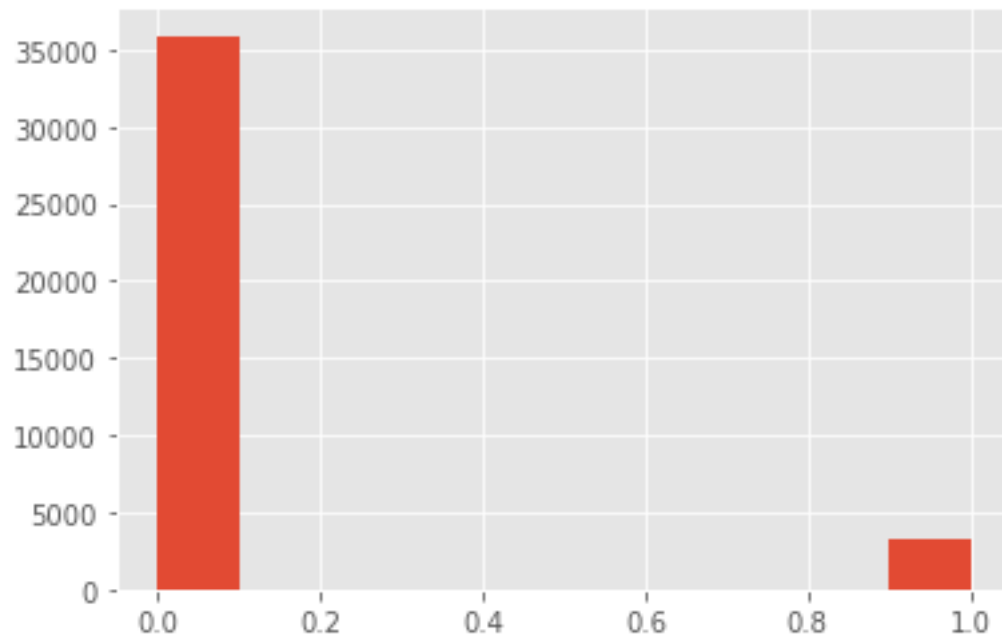
```
[20]: age
16      2717
14      2172
20      2110
4        1611
6        1584
...
116         1
112         1
109         1
108         1
190         1
Length: 131, dtype: int64
```



1.7.11 Anexo 7

```
[21]: A="foreign"  
plt.style.use('ggplot')  
enia['foreign'].hist();  
enia.value_counts(A)
```

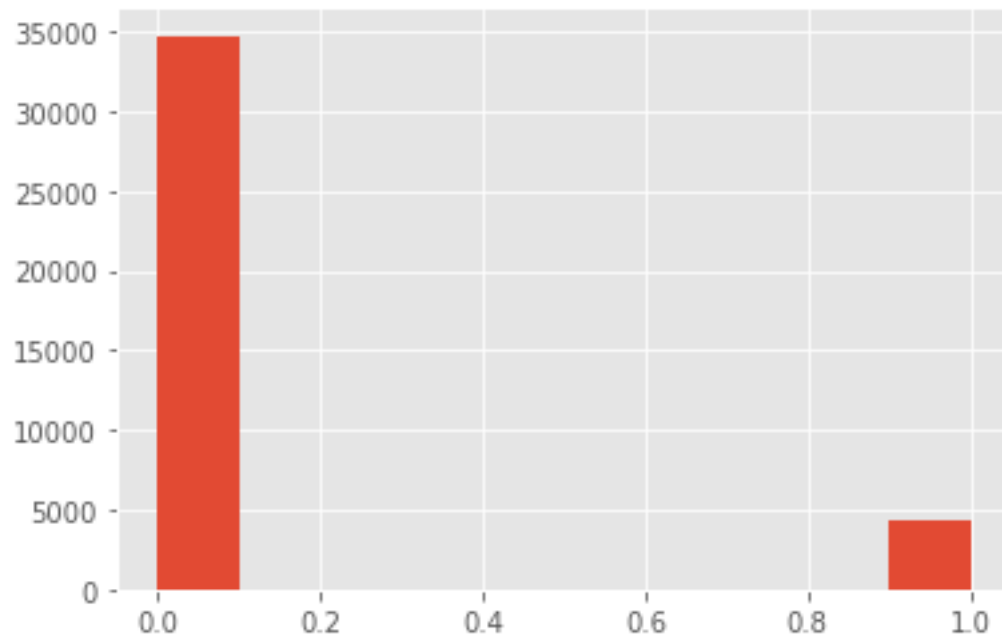
```
[21]: foreign  
0    35903  
1     3201  
dtype: int64
```



1.7.12 Anexo 8

```
[22]: A="export"  
plt.style.use('ggplot')  
enia['export'].hist();  
enia.value_counts(A)
```

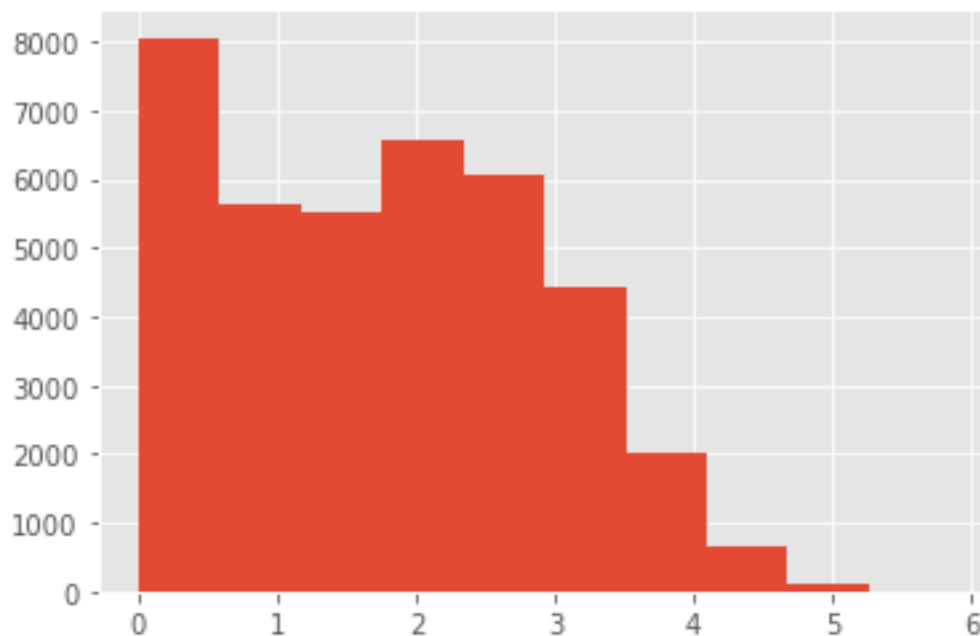
```
[22]: export  
0.0    34756  
1.0     4348  
dtype: int64
```



1.7.13 Anexo 9

```
[23]: A="workers"
plt.style.use('ggplot')
enia['workers'].hist();
enia.value_counts(A)
```

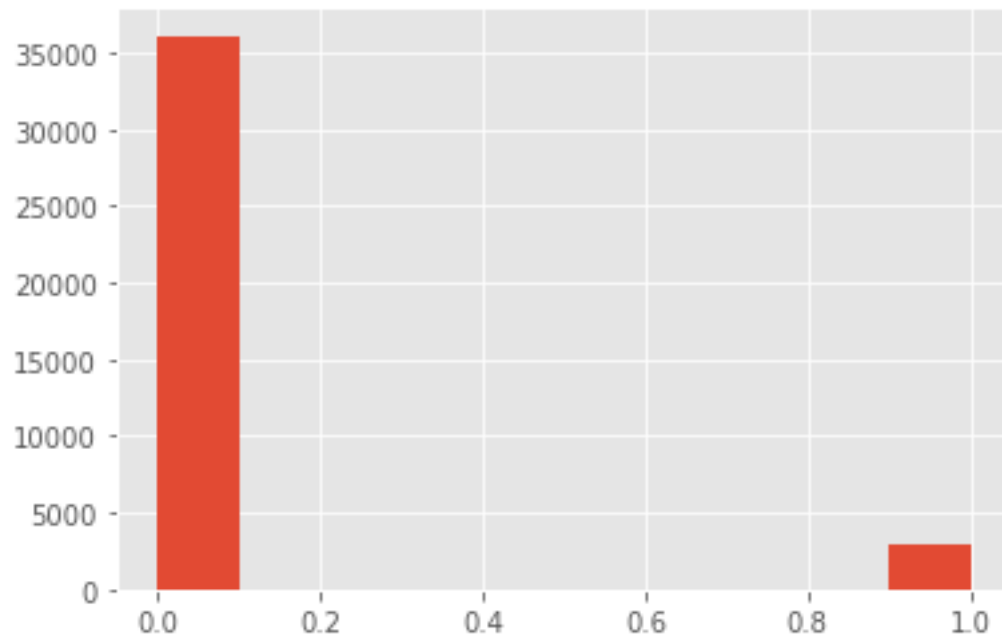
```
[23]: workers
0.000000    5343
0.301030    1568
0.477121    1149
1.113943    1131
1.397940     896
...
3.597256      1
3.597476      1
3.598024      1
3.598243      1
5.845915      1
Length: 4584, dtype: int64
```



1.7.14 Anexo 10

```
[24]: A="fomento"
plt.style.use('ggplot')
enia['fomento'].hist();
enia.value_counts(A)
```

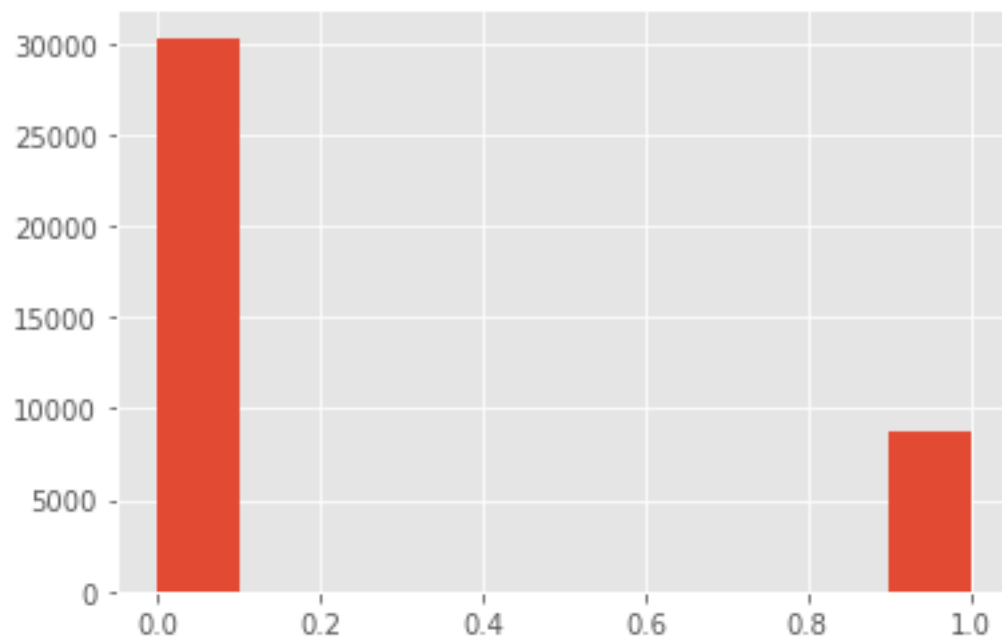
```
[24]: fomento
0    36128
1     2976
dtype: int64
```



1.7.15 Anexo 11

```
[25]: A="iyd"  
plt.style.use('ggplot')  
enia['iyd'].hist();  
enia.value_counts(A)
```

```
[25]: iyd  
0    30310  
1     8794  
dtype: int64
```



1.7.16 Anexo 12

A continuación se va a mostrar la autocorrelación entre variables.

```
[28]: f, ax = plt.subplots(figsize = (14,10))  
sns.heatmap(enia.corr(), annot=True)
```

```
[28]: <AxesSubplot:>
```