

IMPACTO DEL EMPLEO PRECARIO EN LAS MIGRACIONES EXTERIORES

1. Introducción

Nuestro trabajo se basa en estudiar el impacto del número de nuevos contratos y las demandas de empleo por parte de los jóvenes que se realizan cada semestre con las emigraciones. Para analizarlo usaremos los modelos de regresión lineal y reglas de asociación.

2. Objetivos

La hipótesis se basa en confirmar que a mayor demanda de empleo y menor número de nuevos contratos el total de emigraciones aumenta. Para la recogida de datos nos limitaremos en los siguientes grupos

- Edades entre 20 y 40 años divididas en grupos de 5 años
- Contratos indefinidos, obra o servicio y eventuales.

El alcance del trabajo se limita a datos nacionales.

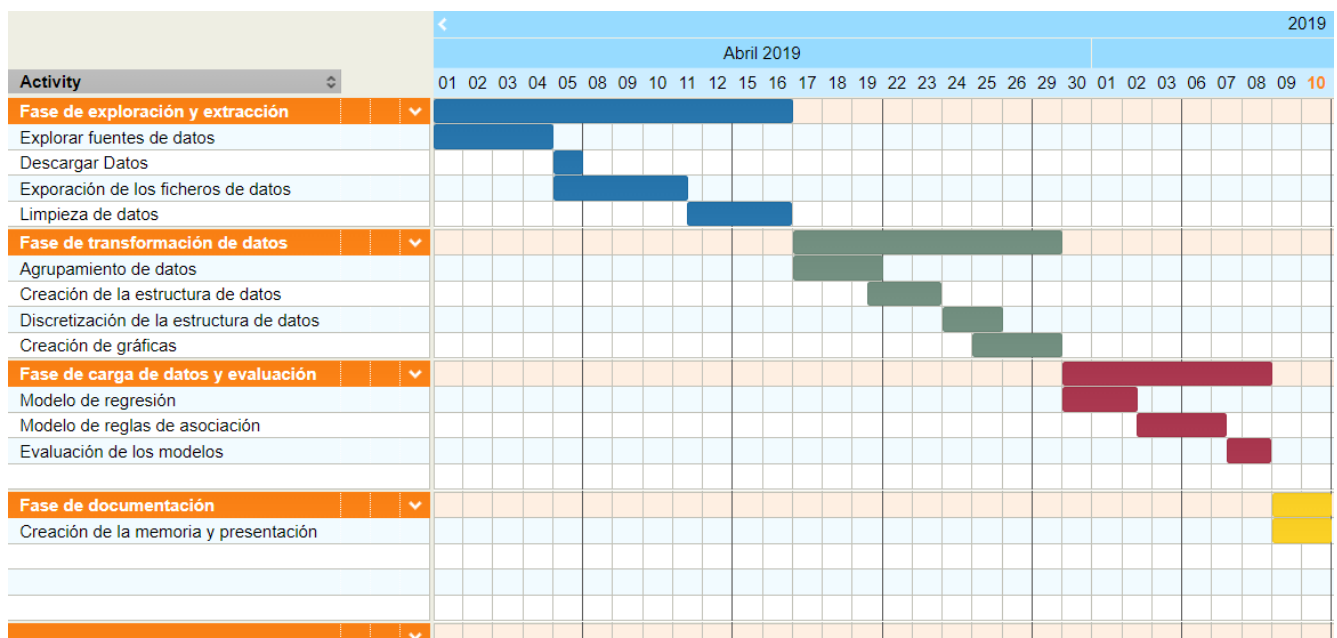


Diagrama de Gantt

3. Metodología

Hemos aplicado la metodología CRISP-DM. La adaptación de las fases a nuestro proyecto es la siguiente:

- Business Understanding

A raíz de pensar la hipótesis e investigando sobre posibles fuentes de datos, nos hemos quedado con la página web del INE para obtener el número de emigraciones y la página web del SEPE para los datos relacionados con contratos y empleo. Nos hemos centrado en el comienzo de la crisis (año 2008 a 2018).

Como los datos de empleo y contratos no tienen la misma estructura temporal que los de emigraciones, tenemos que agruparlos en periodos de 6 meses. Los ficheros tienen formato xls.

- Data Understanding

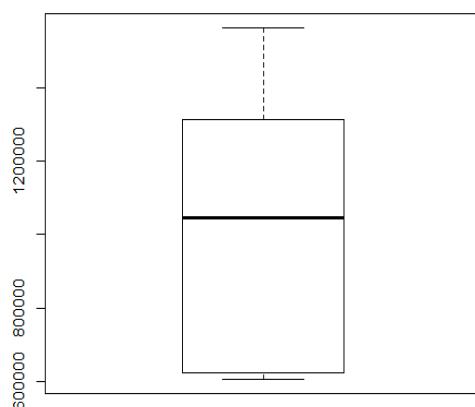
Hemos creado un dataset con todos los ficheros Excel desde el año 2008 al año 2018. Los ficheros se dividen en varias hojas, cada una centrándose en distintas variables relacionadas con empleo o contrato (ocupación, sexo, edad, nivel de estudios...)

Al explorar los datos nos percatamos que la estructura de los Excel no era la misma para todos los años y tuvimos que limitar la investigación hasta el año 2012.

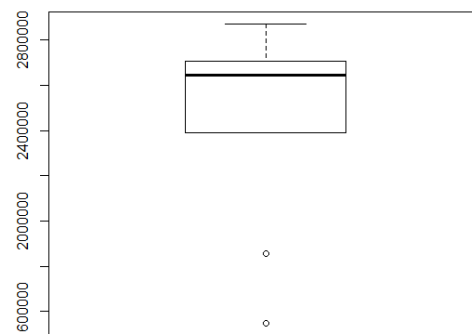
- Data Preparation

Después de explorar todas las hojas, nos quedamos con aquellas que relacionaban los datos con los rangos de edad. Rechazamos variables que no nos interesaban para nuestra hipótesis y creamos un nuevo repositorio con los data.frame que contenían los datos que nos interesaban de cada Excel en bruto.

A partir de ese repositorio creamos un data.frame con los datos discretizados en función de su densidad, divididos por rangos del 20%. El formato de ambos es el mismo, clasificando por semestres y edades.



Boxplot TotalContratos



Boxplot TotalEmpleo

- Modeling

A la hora de aplicar las técnicas de modelado, las dos posibles soluciones que se podían relacionar con nuestra hipótesis eran la regresión y las reglas de asociación.

Para la regresión hemos empleado los datos numéricos, mientras que para las reglas de asociación hemos utilizado los data.frame discretizados.

Los parámetros empleados para la regresión han sido los pares de emigraciones y demandas de empleo, con la finalidad de aproximar la relación de dependencia entre ambas variables. Al evaluar el modelo descubrimos que las variables tenían una relación lineal inversa.

Respecto a la relación de las variables el coeficiente de correlación es de -0.69 y conseguimos un valor de *R-squared* igual a 0.4844, lo que nos indica que según el modelo la mitad de los datos de relaciones dependen de empleo.

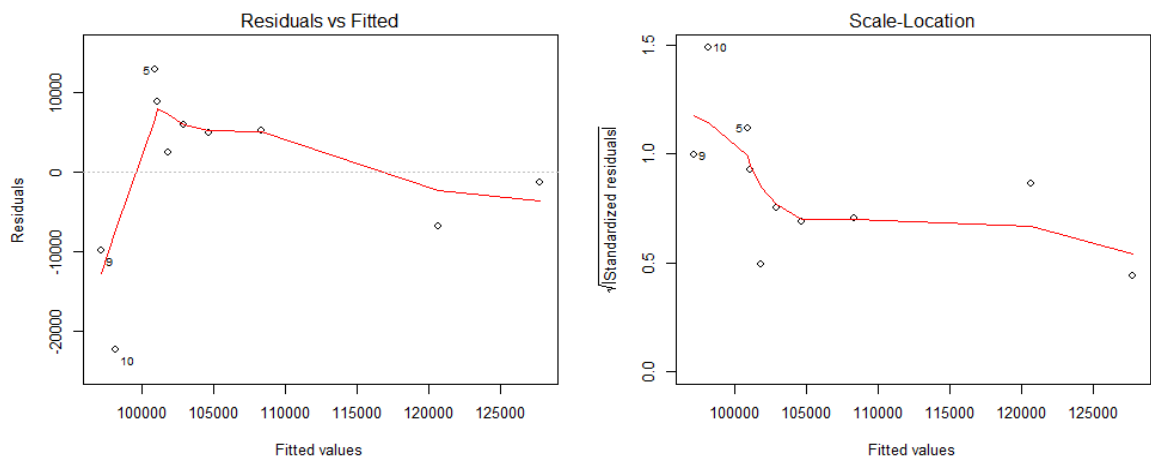
```
Call:
lm(formula = emigraciones$totalEdades ~ empleo$total)

Residuals:
    Min       1Q   Median       3Q      Max
-22211  -5418   3721   5725  12800

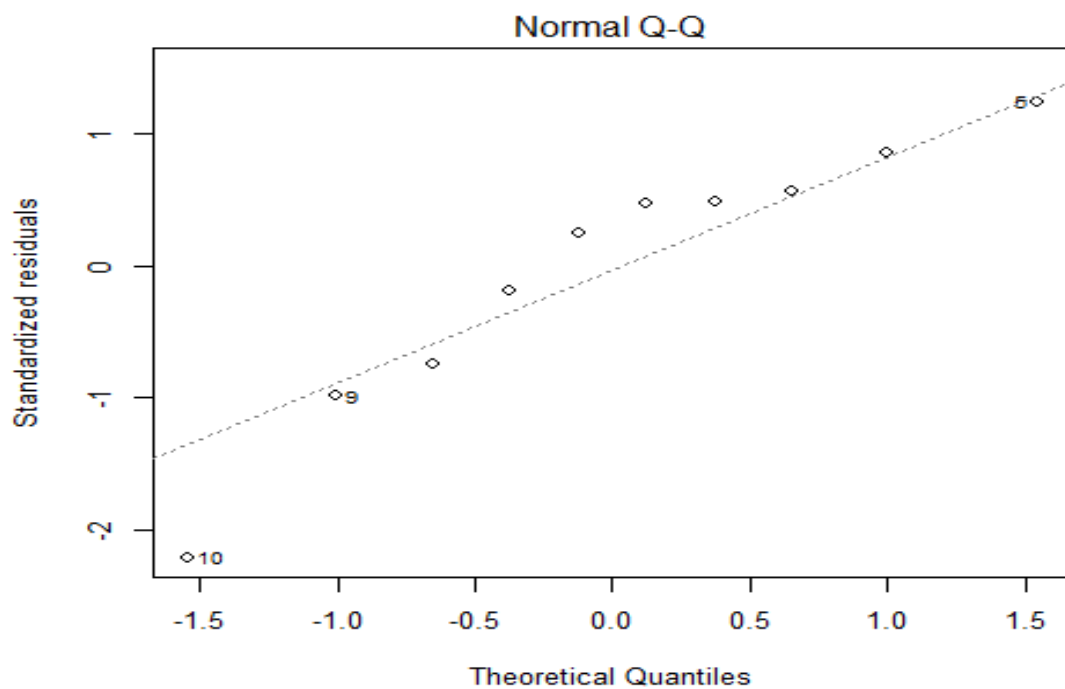
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.637e+05  2.122e+04   7.715 5.67e-05 ***
empleo$total -2.320e-02  8.463e-03  -2.742  0.0254 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11010 on 8 degrees of freedom
Multiple R-squared:  0.4844,    Adjusted R-squared:  0.42
F-statistic: 7.517 on 1 and 8 DF,  p-value: 0.02538
```

Parámetros de la regresión



Gráficas con datos de la regresión



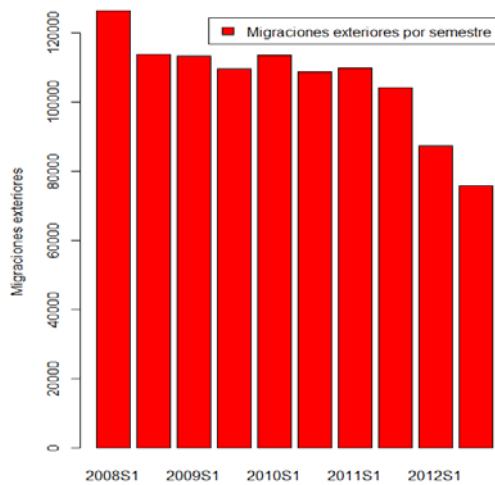
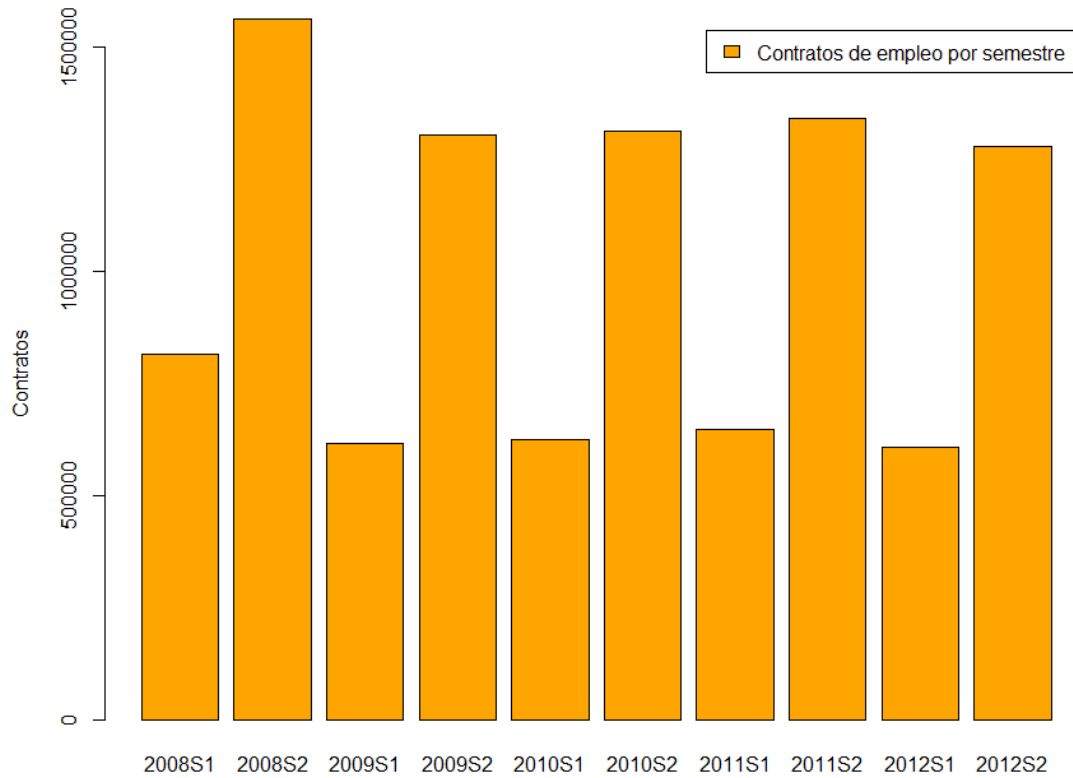
Recta de normalidad de los datos

Para las reglas de asociación, hemos utilizado los datos de empleo y contratos como antecedentes de las reglas y emigraciones como consecuentes. Una vez que obtuvimos las reglas, se filtran para obtener solo las que tienen el consecuente buscado y las hemos ordenados por la confianza en la regla. Finalmente, se han eliminado las reglas redundantes.

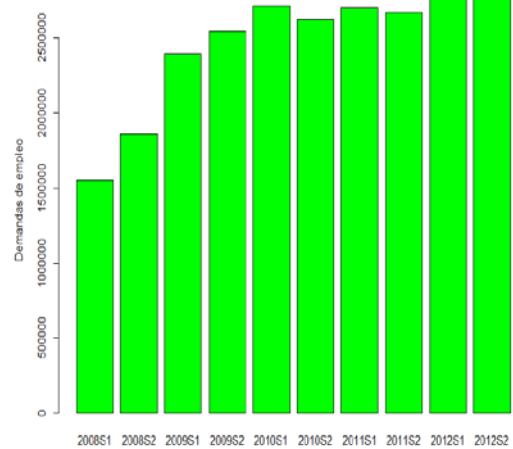
También hemos intentado realizar un árbol de decisión con nuestros datos, pero no hemos conseguido obtener ningún resultado debido a que disponíamos de pocos datos y no se podía clasificar la salida a partir de los contratos ni las demandas de empleo.

- Evaluation

Evaluación de los data.frame: hemos observado que hay mucha diferencia entre los datos de contratos entre ambos semestres, siendo muy bajos los números del primer semestre y bastante altos los del segundo. Respecto a las emigraciones, observamos que tienen una evolución negativa con el paso de los años, mientras que la demanda de empleo aumenta.



Migraciones al exterior



Demandas de empleo

Regresión: entre los coeficientes que resultan de la aplicación de la regresión lineal sobre nuestros datos, el *p-valor* nos indica la validez de nuestro modelo sobre un nivel de confianza elegido (en nuestro caso 0.05). Al ser este (0.02538) menor que el nivel de confianza, confirmamos la fiabilidad del modelo.

Reglas de asociación: hemos identificado que hay una relación oculta entre las dos variables, ya que el nivel de confianza de las reglas es igual a 5. Esto puede deberse a la disparidad de los datos.

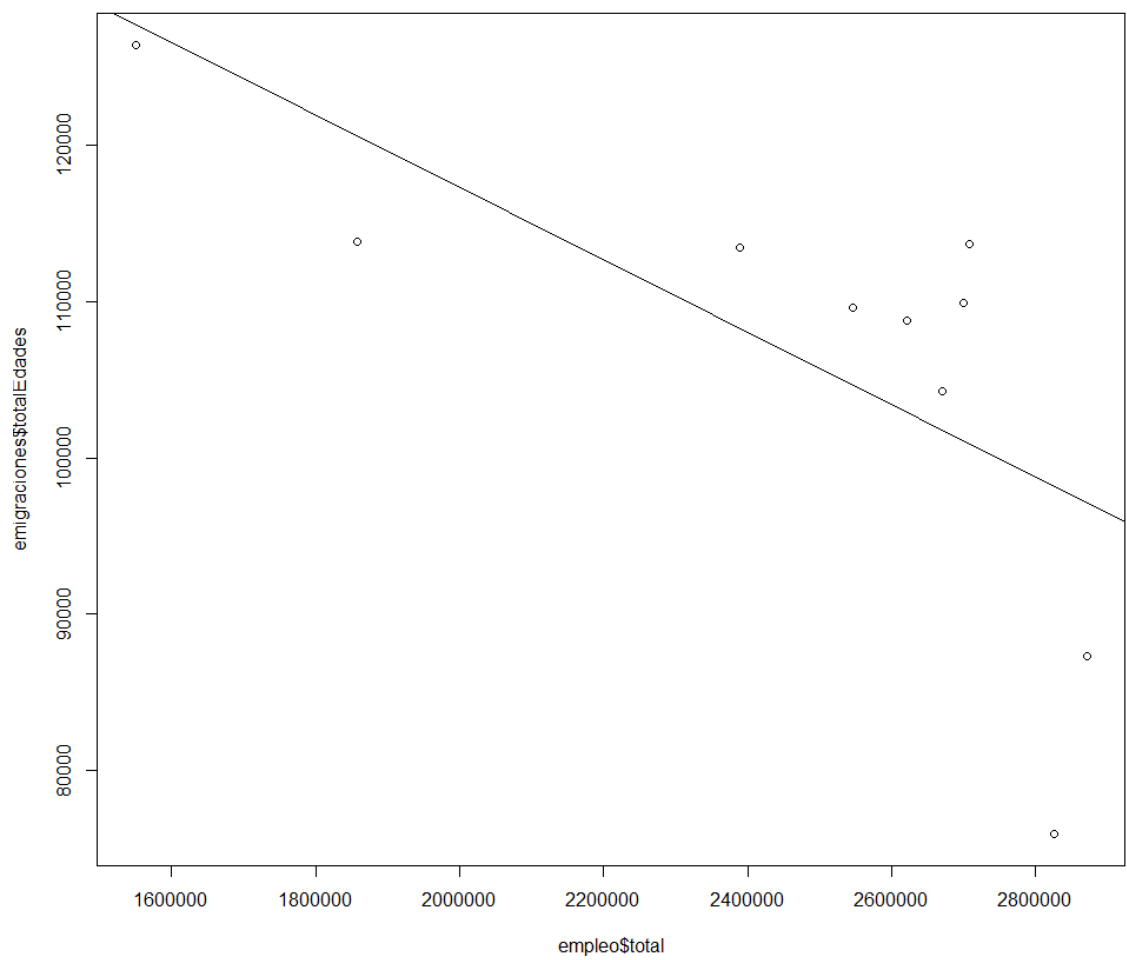
- Deployment

Entregamos esta memoria como documentación del proyecto y realizamos la presentación para la exposición oral.

4. Implementación marcada por las tareas

- Explorar fuentes de datos: como los datos los proporcionan estamentos oficiales (INE y SEPE) en formato Excel, decidimos sacarlos de allí.
- Descargar datos: los datos venían clasificados por meses en el caso de empleo y contratos y por semestres en las emigraciones.
- Exploración de los ficheros de datos: al explorar los ficheros Excel nos dimos cuenta de que la estructura no era común para todos los años, incluso en algún año variaba entre diferentes meses.
- Limpieza de datos: omitimos aquellos datos que no considerábamos relevantes, ya fuera por el tipo de empleo/contrato o por el rango de edad de las solicitudes.
- Agrupamiento de datos: empleos y contratos por semestres para que tuvieran la misma estructura que las emigraciones.
- Creación de la estructura de datos: hemos creado data.frames que contienen los datos de empleos y contratos de cada semestre. El data frame relacionado con el empleo divide las filas por rangos de edad, mientras que el de contratos estructura los datos en filas por rangos de edad y columnas por tipo de contrato.
- Discretización de las estructuras de datos: la división de los datos se ha realizado siguiendo un criterio del 20% para cada valor. Los valores que toman las variables son muy bajo, bajo, normal, alto, muy alto (0% - 20%, 20% - 40%, 40% - 60%, 60% - 80%, 80% - 100%). Se ha creado un data.frame equivalente a los numéricos pero con las variables discretizadas.
- Creación de gráficas.
- Modelo de regresión: hemos cargado los data.frame sin discretizar de empleos y emigraciones. Hemos aplicado el modelo de regresión lineal de R (*función lm()*) para conseguir los parámetros y la recta de la regresión y poder sacar conclusiones.
- Modelo de reglas de asociación: hemos cargado los data.frame discretizados de contratos, empleo y emigraciones y los hemos combinado en otro. Hemos utilizado el paquete *arules* para aplicar el algoritmo *apriori* sobre el data.frame resultante para conseguir las reglas de asociación. A continuación, filtramos las reglas de forma que aparezcan solo aquellas que tienen las emigraciones como consecuente y por último hemos eliminado las reglas redundantes.
- Evaluación de los modelos.
- Creación de la memoria: documentación del proyecto.

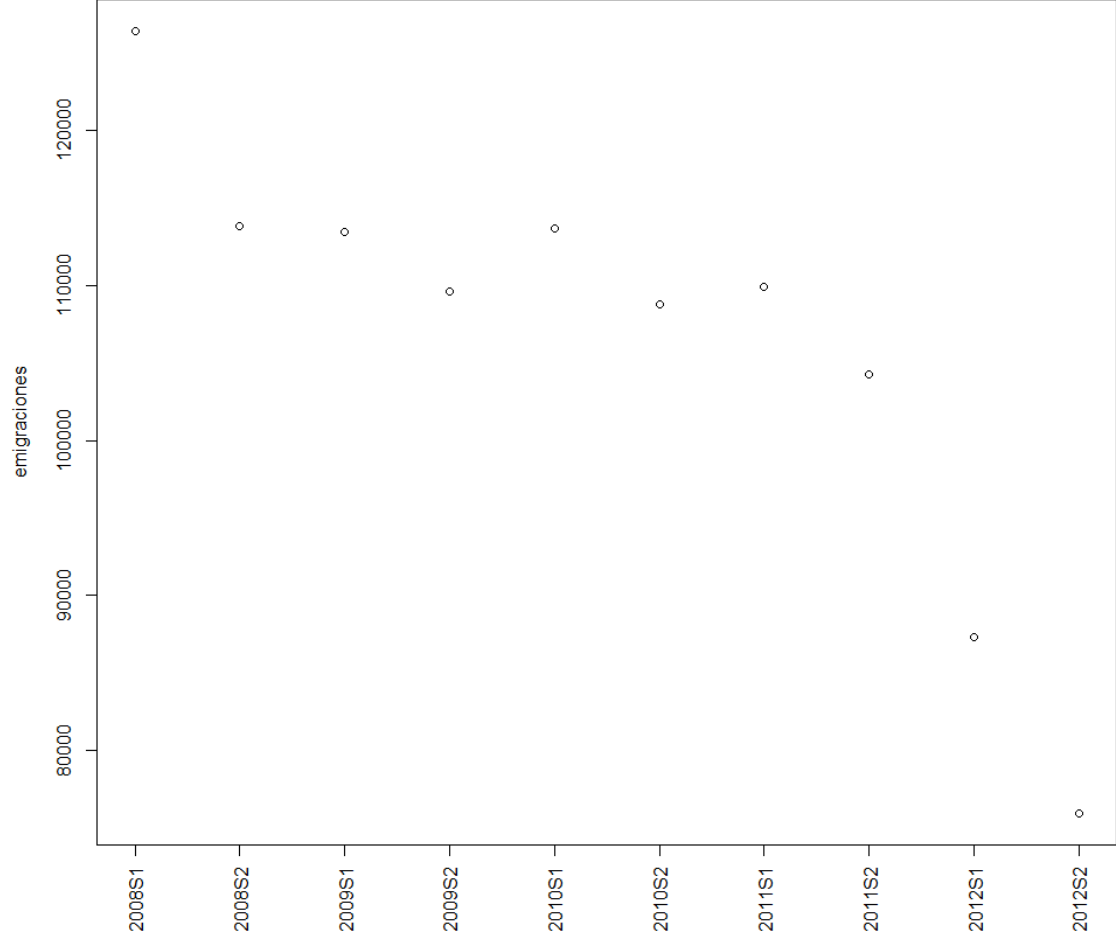
5. Resultados

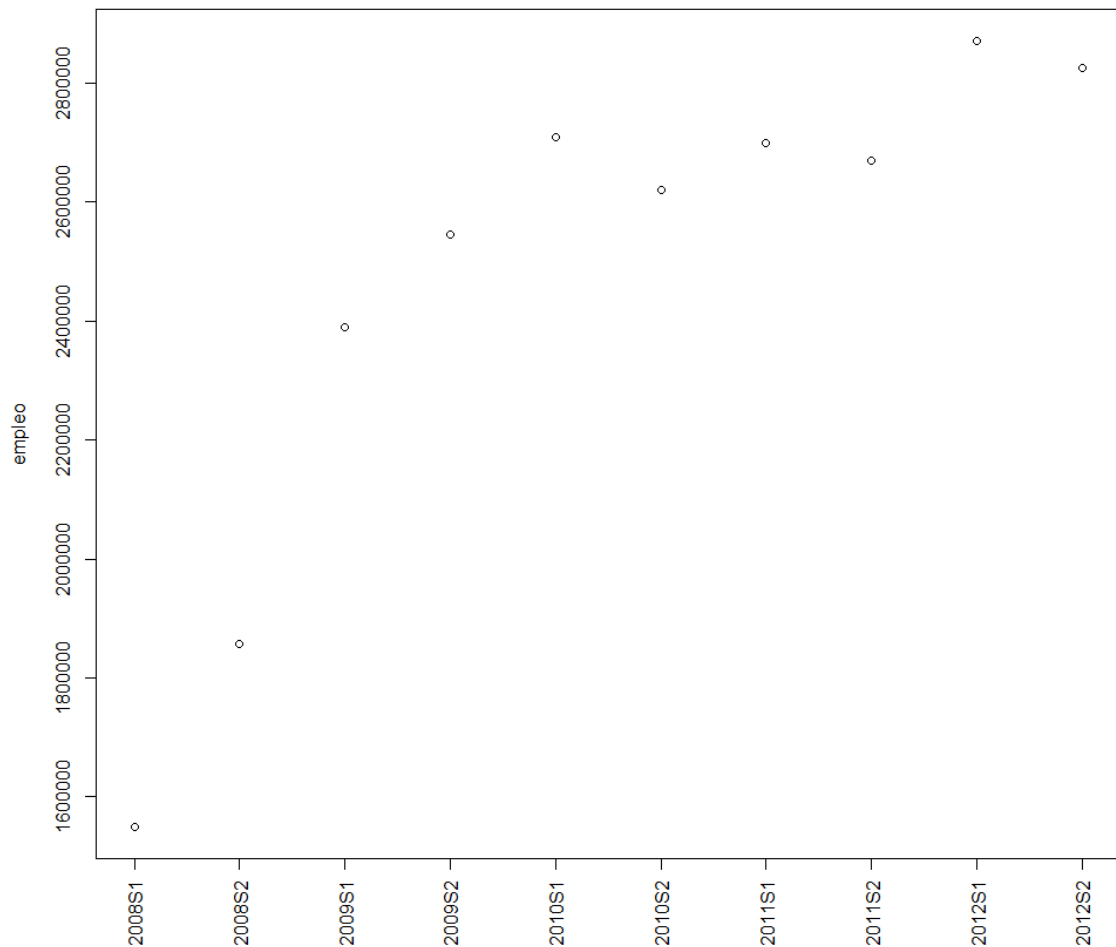


Recta de regresión

lhs	rhs	support	confidence	lift	count
[1] {empleoDiscretizado.empleoTotal=Muy Alto}	=> {emigracionesDiscretizado.emigracionesTotal=Muy bajo}	0.2	1	5	2
[2] {empleoDiscretizado.empleoTotal=Muy bajo}	=> {emigracionesDiscretizado.emigracionesTotal=Muy Alto}	0.2	1	5	2
[3] {empleoDiscretizado.empleoTotal=Normal}	=> {emigracionesDiscretizado.emigracionesTotal=Bajo}	0.2	1	5	2
[4] {contratosDiscretizados.contratosTotal=Alto, empleoDiscretizado.empleoTotal=Bajo}	=> {emigracionesDiscretizado.emigracionesTotal=Normal}	0.1	1	5	1
[5] {contratosDiscretizados.contratosTotal=Muy bajo, empleoDiscretizado.empleoTotal=Bajo}	=> {emigracionesDiscretizado.emigracionesTotal=Alto}	0.1	1	5	1

Reglas de asociación





6. Conclusiones

La relación entre las variables nos demuestra que a mayor número de emigraciones existen menos demandantes de empleo, mientras que los años que se reducen las emigraciones aumenta bastante el número de demandas de empleo en el país.

7. Bibliografía

Fuentes de datos:

- <https://www.ine.es/jaxiT3/Tabla.htm?t=24395&L=0>
- <https://www.sepe.es/HomeSepe/que-es-el-sepe/estadisticas/empleo/estadisticas-nuevas.html>
- <https://www.sepe.es/HomeSepe/que-es-el-sepe/estadisticas/contratos/estadisticas-nuevas.html>

Páginas web consultadas durante el proyecto:

- <http://www.sthda.com/english/articles/2-r/4-xlconnect-read-write-and-manipulate-microsoft-excel-files-from-within-r/>
- <http://altons.github.io/rstats/2015/02/13/quick-intro-to-xlconnect/>
- <http://rfunction.com/archives/1499>
- <http://www.endmemo.com/program/R/gsub.php>
- <https://www.rdocumentation.org/packages/data.table/versions/1.12.2/topics/na.omit.data.table>
- <http://apuntes-r.blogspot.com/2015/07/reglas-de-asociacion.html>
- <https://es.stackoverflow.com/>
- <https://medium.com/datos-y-ciencia/introducci%C3%B3n-a-los-modelos-de-regresi%C3%B3n-en-r-6ef5a4c47a8f>
- http://www.dma.ulpgc.es/profesores/personal/stat/cursor4ULPGC/6g-Data_frames-Listas.html
- <http://www.dma.ulpgc.es/profesores/personal/stat/cursor4ULPGC/9a-graf-Intro.html>
- <https://community.rstudio.com/t/xlconnect-problem/20345>
- <https://stat.ethz.ch/R-manual/R-devel/library/base/html/strtoi.html>
- <https://stackoverflow.com/questions/4931545/converting-string-to-numeric>
- <https://www.rdocumentation.org/packages/utils/versions/3.6.0/topics/type.convert>
- http://rstudio-pubs-static.s3.amazonaws.com/253338_f93c4322ee204a1e993a68553abf6d43.html

Alumnos

- Ernesto Pérez Montalvo
- Federico Sáez Lombán
- Ignacio Vítores Sancho