

Grau en Estadística

Títol: Anàlisi estadística dels "Papers de Panamà"

Autor: Alex Romero Honrubia

Director: Montserrat Guillen Estany

Departament: Econometria, Estadística i Economia aplicada

Convocatòria: Juny 2017

:

Resum

En els últims anys, gràcies a l'avanç imparable de la tecnologia i dels nous mètodes de comunicació, el nombre de filtracions i escàndols que surten a la llum cada any és major. I cada cop, d'àmbits que tradicionalment han estat molt opacs o desconeguts per la majoria de la població. Dintre dels últims escàndols més sonats, trobem el portal *Wikileaks*, conegut per la filtració massiva de documents militars classificats dels EUA i la filtració que dona peu a aquest treball, el *Panama Leaks*, que va destapar centenars de milers de societats registrades a desenes de paradisos fiscals. Exposant les pràctiques utilitzades per importants i influents polítics, cases reials, grans empreses i tot tipus de personalitats mundials per evadir impostos, blanquejar capital i d'altres pràctiques il·legals o poc ètiques. Dintre del *Panama Leak*, hi ha un bufet d'advocats que destaca en la creació d'aquestes societats, el Bufet Mossack Fonseca, font d'informació a partir de la qual es desenvolupa aquest treball. On es realitzarà mitjançant una anàlisi amb el software estadístic R, una comparació entre el comportament anual de les societats espanyoles amb el comportament de la globalitat de les societats del bufet.

PARAULES CLAU

Modelització, evolució anual, societats, paradís fiscal, comparació.

Abstract

In recent years, thanks to the unstoppable advance of technology and new methods of communication, the number of leaks and scandals that come to light is higher and there are increasingly cases of areas that have traditionally been opaque and unknown to most of the population. Within recent scandals, we find portal Wikileaks, known for its massive filtration of classified US military documents and known also for the filtration that gives rise to this work, the Panama Leaks, that uncovered hundreds of thousands of companies registered in dozens of tax havens . Exposing the practices used by important and influential politicians, royal families, main companies and all kinds of world personalities to evade taxes, launder capital and other illegal or

unethical practices. Within the Panama Leak, there is a law firm that excels in the creation of these companies, the Buffet Mossack Fonseca, the source of information from which this work is develop. A statistical analysis will be made with the software R, an annual comparison between the behavior of the Spanish societies and the behavior of worldwide buffet societies’.

KEY WORDS

Modeling, annual evolution, entities, tax haven, comparison.

Classificació AMS

- 62J05 *Linear regression*
- 65D18 *Computer graphics, image analysis and computational geometry*
- 62F03 *Hypothesis testing*
- 97K40 *Descriptive statistics*

ÍNDEX

1	INTRODUCCIÓ	3
1.1	Justificació del treball	4
1.2	Objectius del treball	4
1.3	Estructura del treball	4
1.4	Agraïments	5
2	DESCRIPCIÓ DELS FITXER I LA BASE DE DADES	6
2.1	Programes utilitzats.....	6
2.2	Tipus de fitxers i bases de dades	6
2.2.1	Entities.....	6
2.2.2	<i>Intermediaries</i>	8
2.2.3	<i>Officers</i>	9
2.2.4	<i>Addresses</i>	10
2.2.5	<i>All Edges</i>	11
2.3	Diagrama de la base de dades.....	12
3	ANÀLISI DESCRIPTIVA I COMPARACIÓ GRÀFICA.....	13
3.1	Anàlisi descriptiva de les dades de Mossack Fonseca.....	14
3.1.1	Any de creació de les societats	14
3.1.2	Any de desactivació de la societat	15
3.1.3	Paradisos fiscals.....	16
3.2	Comparació gràfica entre les dues poblacions	17
3.2.1	Evolució del nombre de societat offshore creades i desactivades.....	18
3.2.2	Relació del nombre de societats creades i desactivades	19
3.2.3	Evolució anual del nombre de societats actives diferenciat per Paradís Fiscal ...	20
3.2.4	Nombre de societats creades als diferents paradisos fiscals presents als papers de Panamà.....	22
3.2.5	Evolució anual del total de les societats <i>offshore</i> actives presents als Papers de Panamà	23
4	MODELITZACIÓ I COMPARACIÓ ANALÍTICA.....	25
4.1	Evolució del nombre de societat <i>offshore</i> creades.....	26
4.1.1	Evolució del nombre de societats <i>offshore</i> creades.....	26
4.1.2	Evolució del nombre de societat <i>offshore</i> creades sense l'outlier	28
4.2	Evolució anual del nombre societats desactivades.....	30
4.2.1	Evolució anual del nombre societats desactivades.....	30

4.2.2	Evolució anual del nombre societats desactivades sense l'outlier	32
4.3	Societats actives per any	33
4.4	Temps mitjà de vida de les societats	36
4.5	Comparació de l'evolució anual de societats creades per paradís fiscal	38
4.5.1	Paradís fiscal 1: <i>British Virgin Islands</i>	39
4.5.2	Paradís fiscal 2: Niue	41
4.6	Comparació de l'evolució anual de societats actives per paradís fiscal	42
4.6.1	Paradís fiscal 1: <i>British Virgin Islands</i>	43
4.6.2	Paradís fiscal 2: Niue	45
4.6.3	Paradís fiscal 3: Panamà	47
5	CONCLUSIONS	50
	BIBLIOGRAFIA	52
	ANNEX	53

1 INTRODUCCIÓ

Els Papers de Panamà és el nom que el món periodístic va assignar a una de les majors filtracions de documents fins al moment. En total van ser filtrats 2.6 TB de documents dels servidors del bufet d'advocats Mossack Fonseca, que té com a seu principal Panamà, d'aquí el nom de la filtració. La filtració es va realitzar per una font desconeguda al diari alemany *Süddeutsche Zeitung* i va ser compartida posteriorment amb el *Consortium of Investigative Journalists (ICIJ)*. En aquests documents apareixen centenars de milers de societats *offshore* enregistrades, la gran majoria a països considerats paradisos fiscals, amb els noms dels intermediaris i de les persones que realitzaven la creació de les societats opaques.

El bufet Mossack Fonseca s'encarregava de tota la documentació necessària per a la creació de les societats, de designar els administradors de les mateixes societats i de crear un entramat empresarial. De tal manera, que no es pogués arribar a vincular el compte o la societat *offshore* amb la persona que sol·licitava la seva creació.

Una societat *offshore* és una societat creada a un país diferent del que desenvolupa la seva activitat. És il·legal tenir una societat *offshore*? No, és perfectament lícit, sempre que es declari el capital i les propietats al país on es resideix i es tributi de manera oficial. El perill que tenen aquestes societats *offshore* és que són comunament utilitzades pel desenvolupament de tot tipus d'activitats il·lícites, com pot ser el blanqueig de capitals, l'evasió d'impostos o l'obertura de comptes corrents en paradisos fiscals.

Dintre de l'enorme conjunt de persones que van sortir a la llum involucrades en la creació d'aquestes societats, s'han trobat més de cent cinquanta personalitats exposades públicament, entre les que destaquen: Vladimir Putin, Néstor Kirchner, Lionel Messi o la Infanta Pilar de Borbón. Entre els bancs i societats que han sortit a la llum es pot trobar el BBVA, el Banc Santander i diversos bancs Luxemburguesos.

A dia d'avui encara es segueix ampliant la base de dades dels Papers de Panamà, amb la feina incessant del *Consortium of Investigative Journalists*, que segueix investigant els documents de la filtració.

1.1 Justificació del treball

El motiu de la realització d'aquest treball ve per l'interès en treballar amb grans bases de dades i apropar-me una mica més al món cada cop més present en la nostra societat de el *Big Data*. Per aquest motiu em vaig adreçar a la professora amb la que vaig cursar l'assignatura de Fitxers i Bases de dades, per veure si tenia algun tema interessant que poder treballar. Em va presentar el treball amb els Papers de Panamà.

Degut a l'impacte mediàtic que va suposar aquesta filtració i l'atractiu que representava per a mi aquest tema, en gran part perquè treballo en l'àmbit de la prevenció del blanqueig de capitals i de finançament del terrorisme, em va semblar que treballar amb aquesta base dades i conèixer amb més profunditat la informació que contenia em podria permetre créixer tant com a estadístic com en el meu àmbit professional.

1.2 Objectius del treball

L'objectiu del treball és l'estudi del comportament i tendència de l'evolució anual de les societats que apareixen en els Papers de Panamà. Es realitzarà sobre les societats pertanyents al bufet Mossack Fonseca, i s'analitzaran diverses característiques referents a aquestes societats tals com: el nombre de societats creades anualment, el nombre de societats desactivades anualment, el nombre de societats actives anualment i el temps mitjà de les societats actives.

Aquest estudi és realitzarà comparant entre les societats espanyoles dins la població abans esmentada i la mateixa població, és a dir, serà una comparativa entre la població global i la subpoblació de societats espanyoles. Per veure el comportament de cadascuna de les poblacions i comprovar si es pot considerar igual i, per tant, saber si la influència de les societats espanyoles en el perfil de la població global és rellevant.

1.3 Estructura del treball

La memòria del treball està estructurada en cinc capítols. El primer capítol és una introducció del treball realitzat, els objectius del mateix, els capítols en què està estructurat, amb una petita descripció de cadascun d'aquests capítols i amb els agraïments pertinents.

Al segon capítol es realitza una explicació de les bases de dades utilitzades, amb la descripció i tipus de cada atribut. Així com una descripció dels *softwares* utilitzats per dur a terme el treball.

Al tercer capítol es realitza l'anàlisi descriptiva de les variables principals de les bases de dades i una comparació gràfica de totes dues poblacions, replicant els gràfics trobats a la pàgina font de la qual s'han extret les dades.

Al quart capítol es realitza l'anàlisi en R de totes dues poblacions per a cada característica o punt principal. Es realitzarà la seva comparació mitjançant models de regressió per les dades exposades i un test d'hipòtesi per comparar les variàncies dels models.

Al cinquè capítol es troben les conclusions respecte els objectius fixats a l'inici del treball. A continuació, després de les conclusions es troba la bibliografia i l'annex amb el codi utilitzat per a la realització del treball.

1.4 Agraïments

A la meva tutora Montserrat Guillen, per entendre la meva situació i tenir la paciència i la voluntat de seguir ajudant-me i motivant-me per poder finalitzar el treball de fi de grau.

2 DESCRIPCIÓ DELS FITXER I LA BASE DE DADES

2.1 Programes utilitzats

Per poder dur a terme el treball s'han utilitzat dos *softwares* estadístics així com diferents llenguatges de programació.

El *software* principal ha estat la interfície R-Studio, en el qual s'ha programat amb llenguatge propi R la importació i l'exportació de dades, la creació dels contrastos d'hipòtesi i l'anàlisi dels resultats. Amb el llenguatge SQL, mitjançant el paquet "sqldf", s'ha realitzat el tractament de la base de dades.

L'altre *software* utilitzat ha estat Microsoft Office Excel 2013, on mitjançant les funcions pròpies s'ha procedit a la creació dels gràfics descriptius i comparatius.

2.2 Tipus de fitxers i bases de dades

La base de dades utilitzada ha estat obtinguda de la pàgina propietat de *The International Consortium of Investigative Journalist* (ICIJ). La base de dades consta de cinc datasets independents relacionats per una clau principal, que permet connectar les entitats *offshore* amb els seus intermediaris, directius i adreces informades.

Els datasets, a continuació detallats, que formen la base de dades són: *Entities*, *Officers*, *Intermediaries*, *Addrecess* i *All_edges*.

Aquestes dades han estat filtrades per tractar només les societats vinculades al Bufet d'advocats de Mossak Fonseca.

2.2.1 Entities

El fitxer *Entities* és el dataset que conté la informació principal de les societats *offshore* mencionades a la filtració dels papers de Panamà, està integrada per 319.150 registres i 21 atributs. A continuació s'explica cadascun d'ells:

#	Variable	Type	Format
1	name	Char	
2	original_name	Char	
3	former_name	Char	
4	jurisdiction	Char	
5	jurisdiction_description	Char	
6	company_type	Char	
7	address	Char	
8	internal_id	Num	#
9	incorporation_date	Num	dd-MM-yyyy
10	inactivation_date	Num	dd-MM-yyyy
11	struck_off_date	Num	dd-MM-yyyy
12	dorm_date	Num	dd-MM-yyyy
13	status	Char	
14	service_provider	Char	
15	ibcRUC	Char	
16	country_codes	Char	
17	countries	Char	
18	note	Char	
19	valid_until	Char	
20	node_id	Num	#
21	sourceID	Char	

- *Name*: és el nom de cadascuna de les societats *offshore* pertanyents als Papers de Panamà.
- *Original_name*: és el nom original de les societats *offshore* pertanyents als Papers de Panamà.
- *Former_name*: és el primer nom que van tenir les societats *offshore* pertanyents als Papers de Panamà.
- *Jurisdiction*: sigles del territori.
- *Jurisdiction_description*: territori que té la jurisdicció de la societat *offshore* (paradís fiscal).
- *Company_type*: tipus de societat *offshore* (SA, SL...).
- *Address*: adreça física de l'entitat *offshore*.
- *Internal_id*: identificador del dataset per a cadascuna de les societats *offshore*.

- *Incorporation_date*: data de creació de les societats *offshore* del dataset.
- *Inactivation_date*: data en que les societats *offshore* van deixar d'estar en *good standing*, és a dir, que van deixar de complir amb les seves obligacions de report als organismes centrals i van passar a estar inactiu.
- *Struck_off_date*: data en que la societat ha estat cancel·lada, és a dir, que cap membre d'ella pot actuar de cap manera respecte als actius de la societat.
- *Dorm_date*: data en que la societat va passar a estar inactiva.
- *Status*: estat de la societat *offshore* (actiu, mort...)
- *Service_provider*: proveïdor que dóna els seus serveis per incorporar la societat al territori *offshore*.
- *IbcRUC*.
- *Country_codes*: codificació del país al qual pertany la societat *offshore*.
- *Countries*: país al qual pertany la societat *offshore*.
- *Note*: indicador de que la societat indicada no és una societat *offshore*.
- *Valid_until*.
- *Node_id*: clau principal de la base de dades i identificador principal de les societats *offshore*.
- *SourceID*: origen de la informació de les societats *offshore*. (*Panama papers*, *Offshore leaks*).

2.2.2 Intermediaries

Aquest dataset està format per aquelles persones físiques o jurídiques que feien de pont i registraven les societats en els paradisos fiscals, està integrada per 23.636 registres i 9 atributs. A continuació s'explica cadascun d'ells:

#	Variable	Type	Format
1	name	Char	
2	internal_id	Num	#
3	address	Char	
4	valid_until	Char	
5	country_codes	Char	
6	countries	Char	
7	status	Char	
8	node_id	Num	#
9	sourceID	Char	

- *Name*: nom de l'intermediari.
- *Internal_id*: identificador intern del dataset.
- *Address*: adreça de l'intermediari.
- *Valid_until*: data origen de les dades.
- *Country_codes*: codi d'identificació del país al qual pertany l'intermediari.
- *Countries*: nom del país al qual pertany l'intermediari.
- *Status*: estat en que es troba l'intermediari.
- *Node_id*: identificador i clau principal de la base de dades.
- *Source_id*: font d'origen de les dades.

2.2.3 Officers

Aquest dataset està format per les persones físiques o jurídiques que són executius, directors, advocats o beneficiaris de les societats *offshore*. Aquestes entitats jurídiques poden ser accionistes, oficials o beneficiàries d'altres societats *offshore*, per tant poden aparèixer societats en aquest dataset que formen part també del dataset *Entities*. Està integrat per 345.594 registres i 7 atributs. A continuació s'explica cadascun d'ells:

#	Variable	Type	Format
1	Name	Char	
2	icij_id	Char	
3	valid_until	Char	
4	country_codes	Char	
5	Countries	Char	
6	node_id	Num	#
7	sourceID	Char	

- *Name*: nom de la persona física o jurídica.
- *Icij_id*: identificador intern del dataset.
- *Valid_until*: data origen de les dades.
- *Country_codes*: codi identificació del país al qual pertany l'intermediari.
- *Countries*: nom del país al qual pertany l'intermediari.
- *Node_id*: identificador i clau principal de la base de dades.
- *SourceID*: font d'origen de les dades.

2.2.4 Addresses

Aquest dataset està format per les adreces registrades de les societats *offshore*, i els executius (*officers*), està integrada per 151.054 registres i 7 atributs. A continuació s'explica cadascun d'ells:

#	Variable	Type	Description
1	Address	Char	
2	icij_id	Char	
3	valid_until	Char	
4	country_codes	Char	
5	Countries	Char	
6	node_id	Num	#
7	sourceID	Char	

- *Adress*: adreça de la persona física o jurídica.
- *Icij_id*: identificador intern del dataset.
- *Valid_until*: data origen de les dades.
- *Country_codes*: codi d'identificació del país al qual pertany l'intermediari.
- *Countries*: nom del país al que pertany l'intermediari.
- *Node_id*: identificador i clau principal de la base de dades.
- *SourceID*: font d'origen de les dades.

2.2.5 All Edges

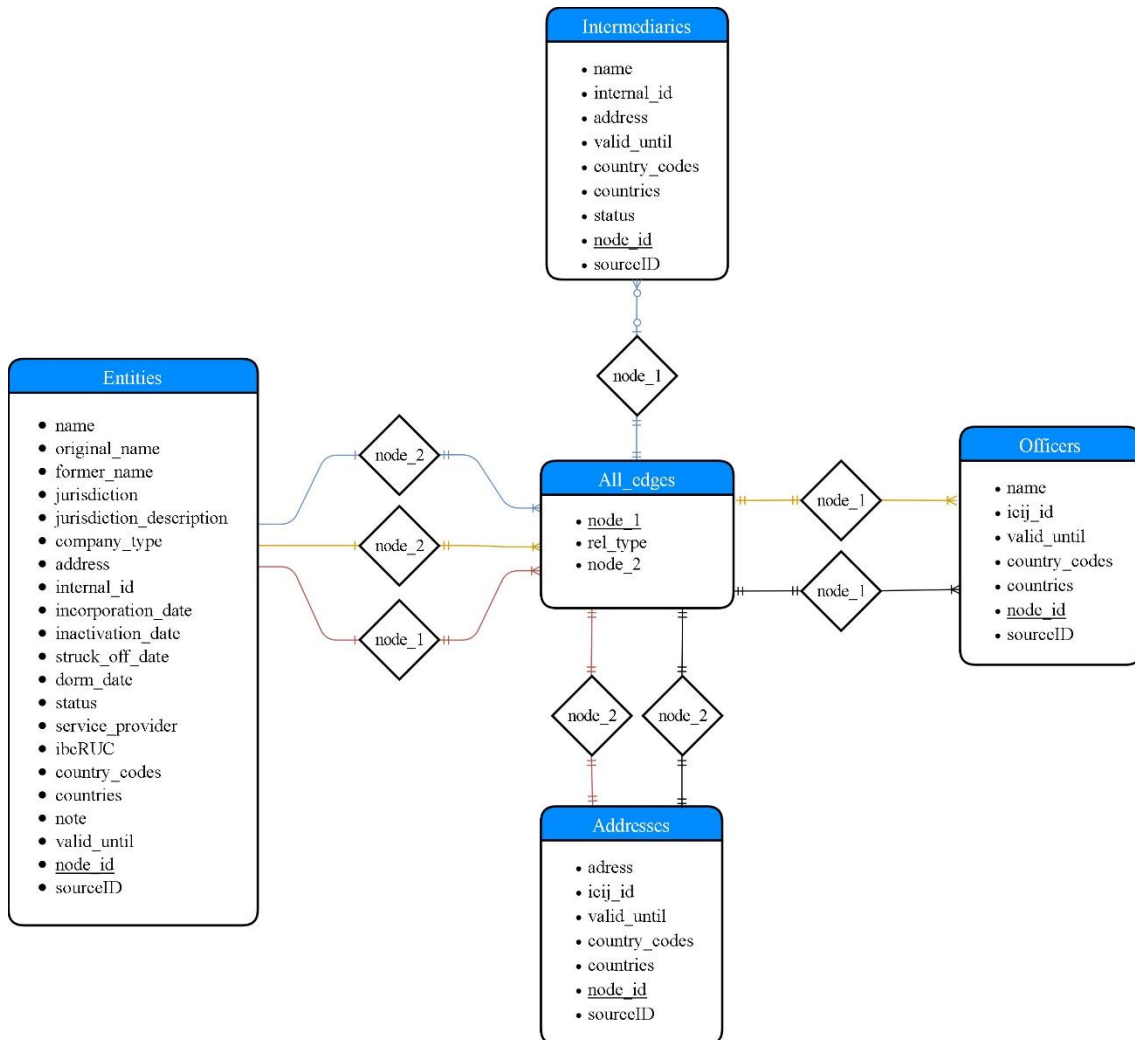
Aquest dataset descriu les relacions que hi ha entre els diferents registres de la resta de datasets. Per exemple, relaciona una entitat *offshore* (amb el seu *node_id* al dataset "Entities") amb el seu advocat (que també té el seu propi *node_id* del dataset "Officers"). També relaciona els dataset "Entites" i "Officers" amb el dataset "Addresses". Està integrada per 1.269.796 registres i 3 atributs. A continuació s'explica cadascun d'ells:

#	Variable	Type	Description
1	node_1	Num	#
2	rel_type	Char	
3	node_2	Num	#

- *Node_1*: identificador *node_id* d'un dels datasets.
- *Rel_type*: tipus de relació que hi ha entre els registres indicats a *node_1* i *node_2*.
- *Node_2*: identificador *node_id* d'un dels datasets.

2.3 Diagrama de la base de dades.

Finalment, per poder representar i apreciar millor la complexitat de la base de dades i les relacions entre els diferents datasets, s'ha realitzat aquest diagrama de flux. Reflexa totes les vinculacions entre els datasets mitjançant el dataset "All_edges".



Gràfic 1. Diagrama de flux de la base de dades. Dades "Panama Papers". Font: Elaboració pròpia.

3 ANÀLISI DESCRIPTIVA I COMPARACIÓ GRÀFICA

En aquest apartat es tracta la creació dels nous datasets a partir dels principals, per poder portar a terme una primera comparació gràfica del comportament de les societats *offshore* espanyoles amb la resta del món. Pertanyents totes dues poblacions a les dades del buffet Mossack Fonseca.

Per a la creació dels nous datasets, s'ha procedit mitjançant el paquet "sqldf", a extreure del dataset "Entities" tots aquells registres que contenen a l'atribut "Countries" la cadena de caràcters Spain. S'ha obtingut com a resultat un nou dataset anomenat "spain". A partir de la creació del dataset "spain" i mitjançant el dataset "all_edges", s'han creat els datasets "spain_offi", amb les dades del dataset "officers" de les societats espanyoles; "spain_inter" amb els intermediaris; "spain_adre" amb les adreces conegudes de les societats i "spain_adre2" amb les adreces conegudes dels *officers* d'aquestes societats. Totes les observacions menys una provenen del buffet Mossack Fonseca, aquesta ha estat eliminada del dataset "spain".

S'ha realitzat també el mateix procediment amb les dades globals per tots aquells registres amb la cadena de caràcters Mossack Fonseca a l'atribut "service_provider" del dataset "Entities", obtenint els datasets "fonseca", "fonseca_inter", "fonseca_offi", "fonseca_adre" i "fonseca_adre2" respectivament.

Per poder procedir a la comparació del comportament de la globalitat de societats *offshore* informades en la base de dades amb el comportament de les societats espanyoles, s'han replicat una sèrie de gràfics que apareixen a la pàgina de *The International Consortium of Investigative Journalists* d'on s'ha obtingut la base de dades.

Per realitzar aquests gràfics comparatius s'ha hagut de crear noves variables que permetessin resumir de manera més concisa la informació que apareix als datasets principals. Aquestes noves variables són:

- Nombre de societats creades en funció de l'any.
- Nombre de societats desactivades en funció de l'any.
- Nombre de societats actives en funció de l'any.
- Nombre de societats actives en funció de l'any i el paradís fiscal.
- Nombre total de societats creades en funció del paradís fiscal.

- Temps mitjà de vida de les societats en funció de l'any de creació.

Els gràfics s'han realitzat amb el *software* Excel, degut a la intenció de divulgar-los públicament. Ja que són molt més visuals i és molt més fàcil poder comparar-los amb els presents a la pàgina d'origen dels Papers de Panamà.

3.1 Anàlisi descriptiva de les dades de Mossack Fonseca

L'anàlisi descriptiva que s'ha realitzat sobre les variables principals dels cinc datasets que comprenen la base de dades dels Papers de Panamà.

3.1.1 Any de creació de les societats

S'ha realitzat una anàlisi descriptiva amb el software R i pel què fa a la variable "incorporation_date" s'han obtingut els següents resultats:

MOSSACK FONSECA	
nbr.val	213599.00
nbr.na	35.00
min	1936.00
max	2015.00
range	79.00
sum	427775523.00
median	2004.00
mean	2002.70
var	55.88
std.dev	7.47

El nombre total d'observacions, és a dir de societats registrades, és de 213.634, amb un total de 34 dades mancants, és a dir, amb 34 societats sense data de creació informada. La primera societat informada va estar creada al 1936 i l'última a l'any de creació de la base de dades, 2015. La mitjana d'any de creació és el 2002 i la mediana és el 2004, que ens indica que hi ha dades més extremes a la primera meitat de les dades que a la segona, fet totalment coherent degut a que la data màxima és 2015, mentre que no hi ha data mínima. Finalment tenim una desviació tipus de 7,47 anys dins la població global.

ESPANYA

nbr.val	1166.00
nbr.null	0.00
min	1969.00
max	2015.00
range	46.00
sum	2327667.00
median	1995.00
mean	1996.28
var	65.29
std.dev	8.08

Respecte al nombre total d'observacions de les societats registrades espanyoles veiem que és de 1.166. Amb la primera societat creada al 1969 i l'última informada a la base de dades a l'any 2015. La mitjana de l'any de creació és 1996 i la mediana 1995, indiquen certa simetria respecte el nombre d'observacions per cada meitat, com es pot veure per la diferència en anys respecte la dada de tendència central. La desviació tipus de la població de societats espanyoles és de 8,08 anys.

3.1.2 Any de desactivació de la societat

S'ha realitzat una anàlisi descriptiva amb el *software* R pel què fa a la variable "inactivation_date" i s'han obtingut els següents resultats:

MOSSACK FONSECA	
nbr.val	144750.00
nbr.na	68884.00
min	1967.00
max	2015.00
range	85.00
sum	290366904.00
median	2007.00
mean	2005.98
var	41.18
std.dev	6.41

El nombre total d'observacions és, òbviament, el mateix que al punt anterior amb la diferència que hi ha 68.884 societats sense data de desactivació informada, és a dir, que a data 2015 encara es troben actives. La primera societat desactivada va estar creada al 1967, tot i que a la base de dades hi ha una dada errònia que indica que la primera va ser al 1930, i l'última a 2015. La mitjana d'any de desactivació és el 2005 i la mediana és el 2007. Finalment tenim una desviació tipus de 6,41 anys dins la població global.

	ESPANYA	
nbr.val		1034.00
nbr.na		132.00
min		1984.00
max		2015.00
range		31.00
sum	2069126.00	
median		2001.00
mean		2001.08
var		44.93
std.dev		6.70

Respecte al nombre total d'observacions de les societats desactivades espanyoles és, igual que al punt anterior, de 1.166. Amb la variació de que hi ha 132 societats que encara segueixen actives. La primera societat desactivada va ser al 1984 i l'última informada a la base de dades a l'any 2015. La mitjana de l'any de desactivació és 2001 i la mediana 2001 també. La desviació tipus de la població de societats espanyoles és de 6,70 anys.

3.1.3 Paradisos fiscals

En aquest subapartat, s'ha realitzat una anàlisi descriptiva amb el *software* R pel que fa a la variable "inactivation_date". Obtenint els següents resultats:

MOSSACK FONSECA	
PARADÍS FISCAL	Nº SOCIETATS
British Virgin Islands	113648
Panama	48325
Bahamas	15915
Seychelles	15182
Niue	9611
Samoa	5307
British Anguilla	3253
Nevada	1260
Hong Kong	452
United Kingdom	148
Belize	130
Costa Rica	78
Cyprus	76
Uruguay	52
New Zealand	47
Jersey	39
Wyoming	37
Malta	28
Isle Of Man	8
Ras Al Khaimah	2
Singapore	1

El nombre total de paradisos fiscals mencionats a la població global és de 21, més de la meitat les societats han estat creades al paradís *British Virgin Islands*, amb una diferència de diverses desenes de milers de societats entre aquest i el segon paradís fiscal i que dóna nom a la base de dades Panamà. Dels vint-i-un paradisos fiscals, tretze tenen enregistrades menys de mil societats, “*Singapore*”, “*Ras Al Khaimah*” i “*Isle of Man*” destaquen amb només una, dues i vuit societats enregistrades, respectivament.

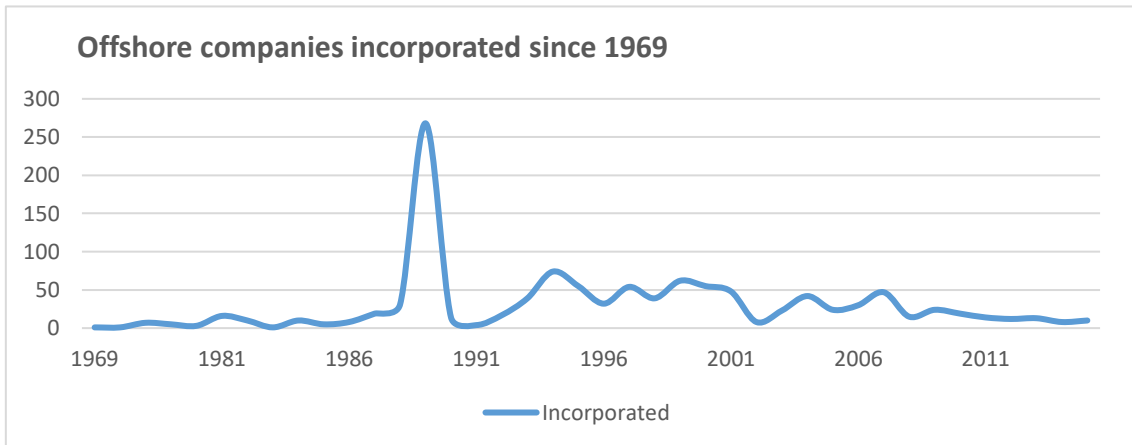
ESPANYA	
PARADÍS FISCAL	Nº SOCIETATS
Panama	540
British Virgin Islands	186
Niue	186
Bahamas	163
Samoa	38
Seychelles	38
Nevada	8
British Anguilla	3
Uruguay	3
Jersey	1

Respecte al nombre total de paradisos fiscals on s’han registrat les societats espanyoles, és mes reduït, i fa un total de 10. En aquesta població però, la meitat de les societats han estat creades al paradís Panamà, amb una diferència de diversos centenars de societats entre aquest i el segon paradís fiscal més utilitzat, les *British Virgin Islands*. Dels 10 paradisos fiscals, 6 tenen enregistrades menys de cent societats, destaquen “*Singapore*” amb una, “*Uruguay*” i “*British Anguilla*” amb tres i “*Nevada*” amb vuit.

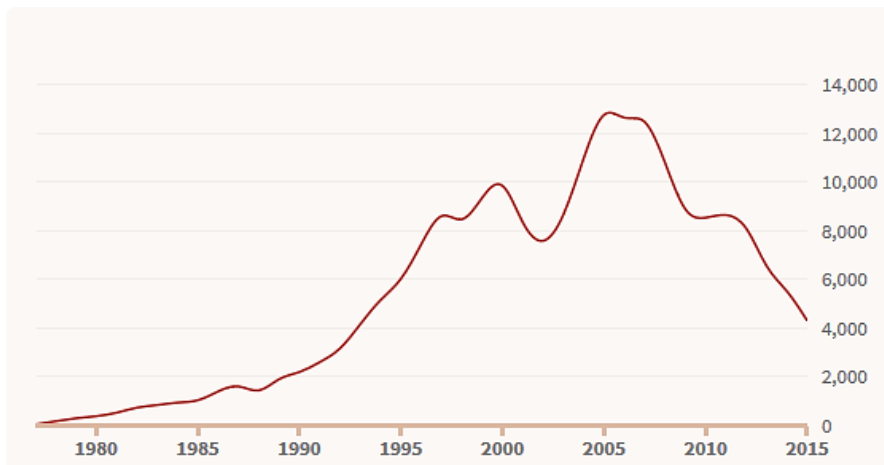
3.2 Comparació gràfica entre les dues poblacions

A continuació s’ha realitzat una primera anàlisi gràfica comparativa entre totes dues poblacions.

3.2.1 Evolució del nombre de societats offshore creades i desactivades

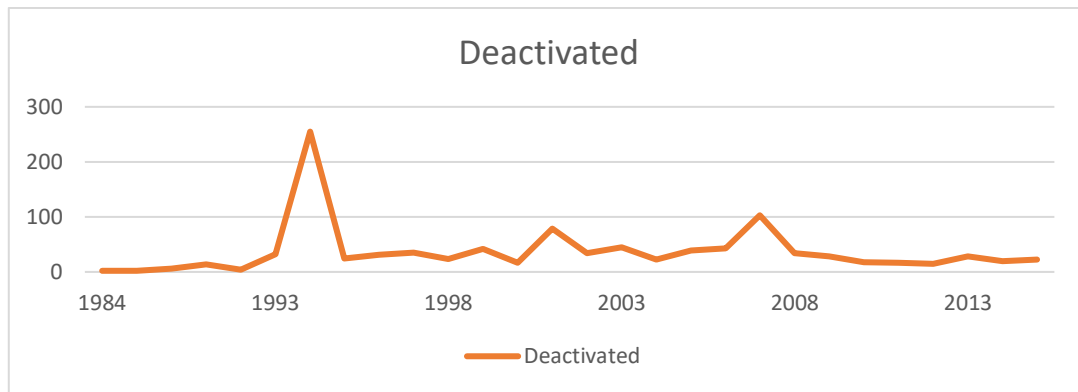


Gràfic 2. Creació de societats offshore. Dades per a SPAIN. Font: Elaboració pròpia.



Gràfic 3. Creació de societats offshore. Dades globals. Font: Web dels Panama Papers.

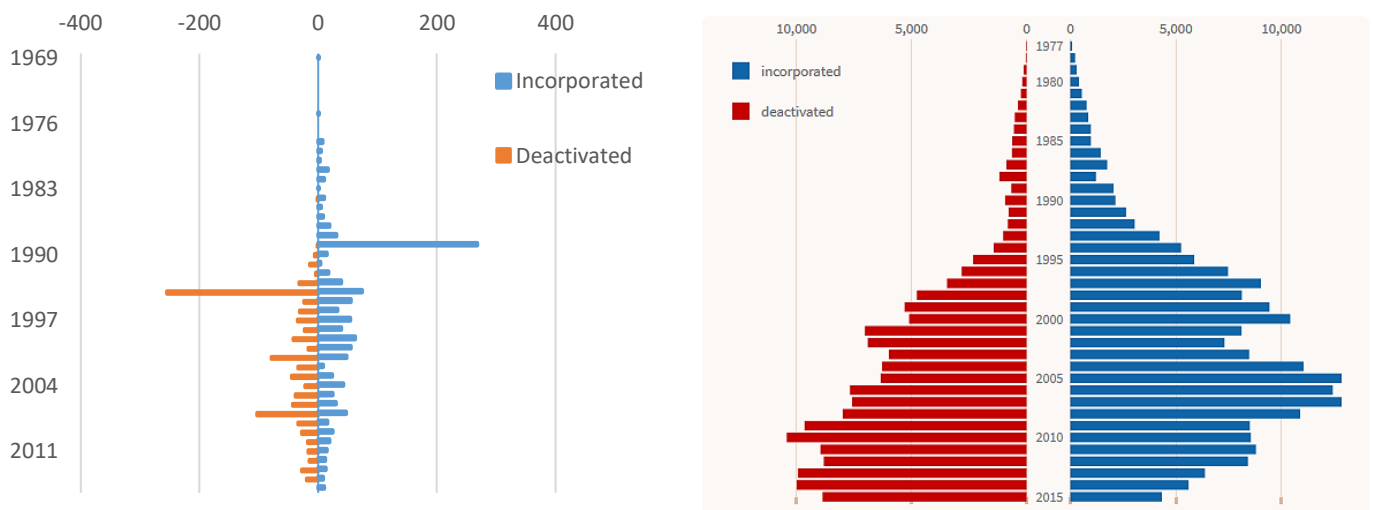
Al primer gràfic es pot observar l'evolució del nombre de societats que s'han anat creant als diversos paradisos fiscals al llarg dels anys, el més destacable és el pic de societats que són creades al 1989 (268) i que als següents anys cau fins a la creació de només 4 noves societats al 1992. A partir de 1993 comença a créixer i es manté gairebé constant durant 17 anys fins al 2008 que comença a decreixer significativament. El segon gràfic representa també la creació anual de noves societats *offshore* però a escala global. Tot i que sembla molt dispar, a causa del pic de 1989 que no ens permet valorar bé gràficament el comportament de la mostra, es pot apreciar certa similitud: el creixement des del començament dels 90, la caiguda al 2002, el nou creixement fins al 2007 i la caiguda constant fins al 2015.



Gràfic 4. Inactivació de societats offshore. Dades per a SPAIN. Font: Elaboració pròpia.

De manera anàloga, pel que fa a les societats desactivades, es pot observar un gran creixement d'inactivació de les societats al 1994. A partir d'aquest d'any el nombre anual de societats que passen a estar en estat inactives es manté constant i inferior al de societats creades, amb excepció de dos nous pics, al 2001 i 2007, on torna a haver-hi un elevat nombre de societats que passen a estar inactives, tots dos un any abans de la caiguda del nombre de societats creades.

3.2.2 Relació del nombre de societats creades i desactivades

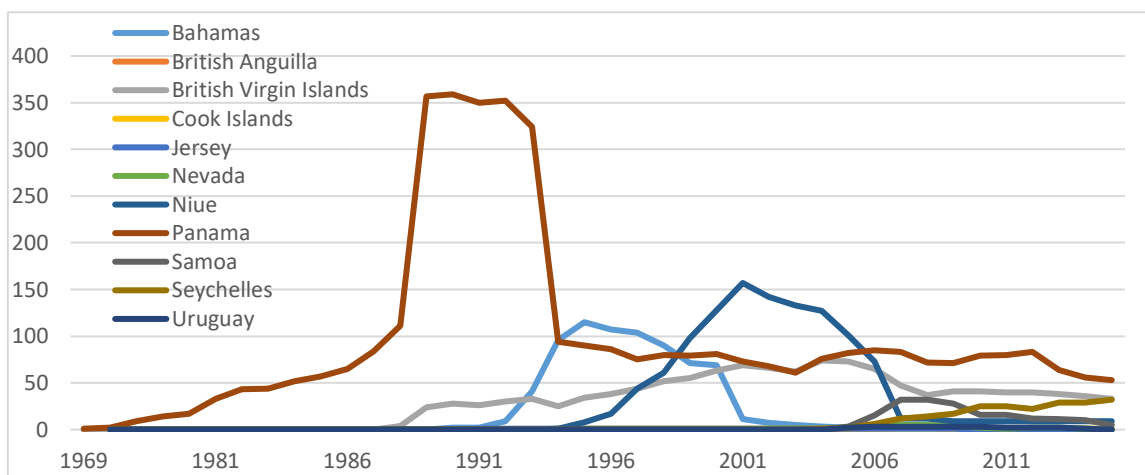


Gràfic 5. Data de creació de societats offshore i piràmide d'edat. Dades SPAIN (esquerra) i dades globals (dreta). Font: Elaboració pròpia i Web dels Panama Papers.

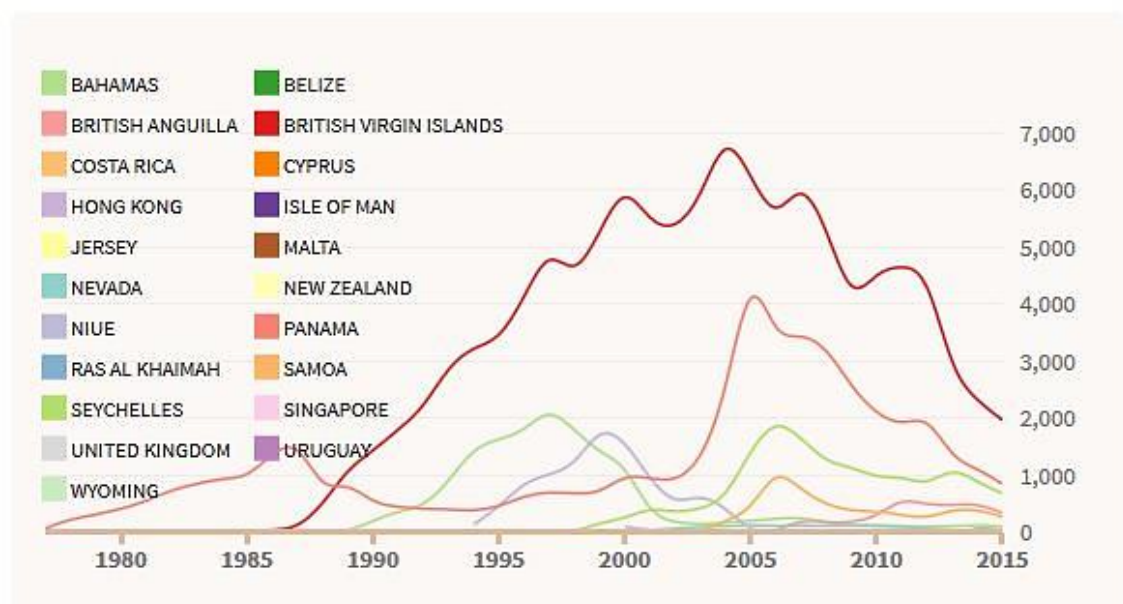
En aquest gràfic de tipus piramidal, es veu clarament la relació entre el nombre de societats *offshore* creades anualment amb el nombre de societats que passen a estar inactives. Igual que s'ha explicat en els gràfics de l'apartat anterior, es pot observar la

relació entre el gran nombre de societats desactivades en 2001 i 2007 i la posterior caiguda de la creació de noves societats *offshore* al 2003 i 2008 i la similitud entre el comportament de la creació de societats *offshore* espanyoles i el comportament global. Aquest fet no es dona amb les societats que passen a estar desactivades. Al gràfic global es pot apreciar com a partir del 2007, a mesura que va disminuint el nombre de societats creades, va augmentant el nombre de societats que passen a estar inactives, mentre que al gràfic de les societats espanyoles a partir de 2007, tot i que el nombre de societats inactives és superior a la creació de noves, totes dues tipologies van disminuint.

3.2.3 Evolució anual del nombre de societats actives diferenciat per Paradís Fiscal



Gràfic 6. Societats Actives segons paradís fiscal. Dades per a SPAIN. Font: Elaboració pròpia.



Gràfic 7. Societats Actives segons paradís fiscal. Dades globals. Font: Web dels Panama Papers.

Aquests gràfics reflecteixen l'evolució anual del nombre de societats *offshore* actives en funció del paradís fiscal. El desglossament del comportament de les societats espanyoles ens permet aprofundir i veure els motius del comportament general analitzat en els apartats anteriors.

Destaquen quatre fets que marquen l'evolució general:

- i. El recurrentment esmentat augment del nombre de societats *offshore* al 1989 i el gran nombre de societats que van ser desactivades al 1994, no va ser un comportament general en tots els paradisos fiscals sinó que es va donar només en les societats *offshore* de Panamà.
- ii. Igual que en les societats panamenyes, al 1994 es dispara, tot i que en menor escala que al Panamà al 1989, la creació de noves societats a les Bahames que passen a la inactivitat al 2001, novament es tracta d'un paradís fiscal concret i no d'un comportament generalitzat.
- iii. De la mateixa manera que als dos paradisos fiscals comentats, també trobem una forta crescuda de les societats *offshore* creades a Niue amb el seu pic de societats actives al 2001 i com passen a desactivar-se gairebé la gran majoria entre 2005 i 2007. Com en els dos punts anteriors es tracta d'un paradís fiscal concret i no d'un comportament generalitzat.
- iv. En els anys posteriors al 2007, en gairebé tots els paradisos fiscals va minvant el nombre de societats *offshore* registrades. Fins i tot a Panamà, que després del succés de 1994 ha mantingut un nombre constant. Només a Seychelles va augmentant molt lentament.

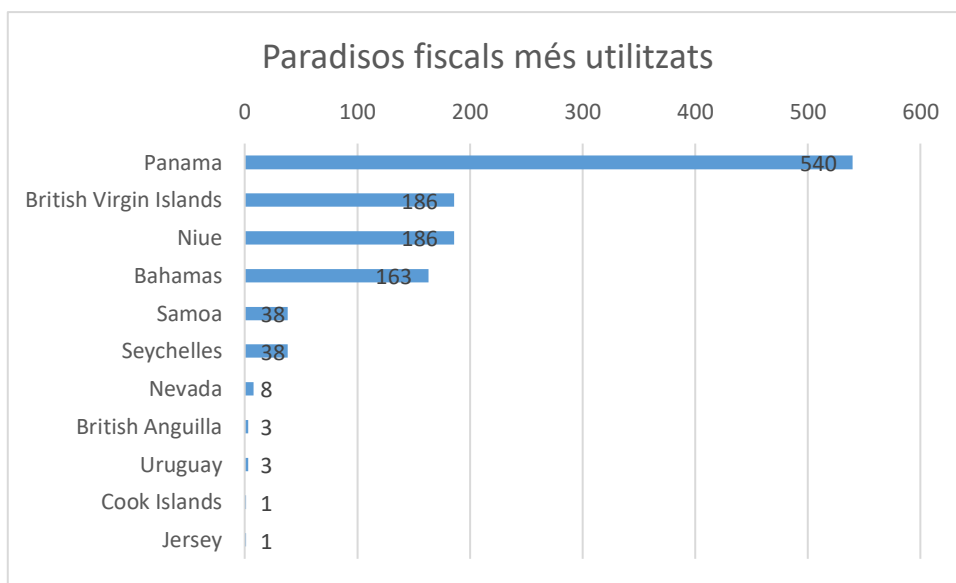
Per poder procedir a la comparativa amb el segon gràfic és necessari dividir-ho en dos subperíodes. El primer subperíode va des de l'inici de les dades fins al 2005 i el segon des del 2005 fins a la data fi de les dades.

Durant el primer període, s'observa com el comportament del nombre de societats actives a Panamà, Niue i les Bahames pateixen una forta pujada seguida d'una clara davallada. Aquest període concorda amb els registrats amb les societats *offshore* espanyoles.

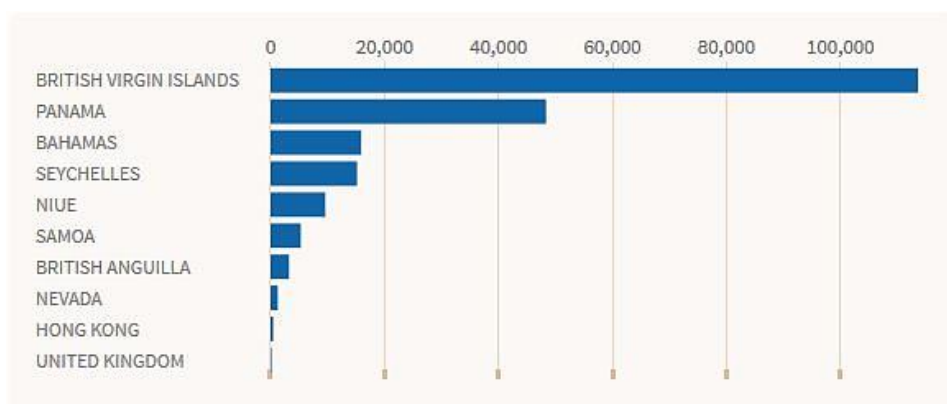
Durant el segon període, el nombre de societats actives es dispara en gairebé tots els paradisos fiscals i comença a caure de manera progressiva, al contrari que en el comportament de les societats espanyoles, on a partir del 2006 cauen totes.

D'altra banda hi ha un paradís fiscal que té un comportament bastant dispar amb el de les societats espanyoles, no en el període d'activitat sinó en la magnitud del nombre de societats *offshore* actives. S'observa clarament que com a escala global, les *British Virgin Islands* són el paradís fiscal amb major nombre de societats *offshore* actives al llarg de les dècades, mentre que a les societats *offshore* espanyoles tenen una proporció molt inferior al què hauria de ser si es comportessin de la mateixa manera.

3.2.4 Nombre de societats creades als diferents paradisos fiscals presents als papers de Panamá



Gràfic 8. Paradisos fiscals més utilitzats. Dades per a SPAIN. Font: Elaboració pròpia.



Gràfic 9. Paradisos fiscals més utilitzats. Dades globals. Font: Web dels Panama Papers.

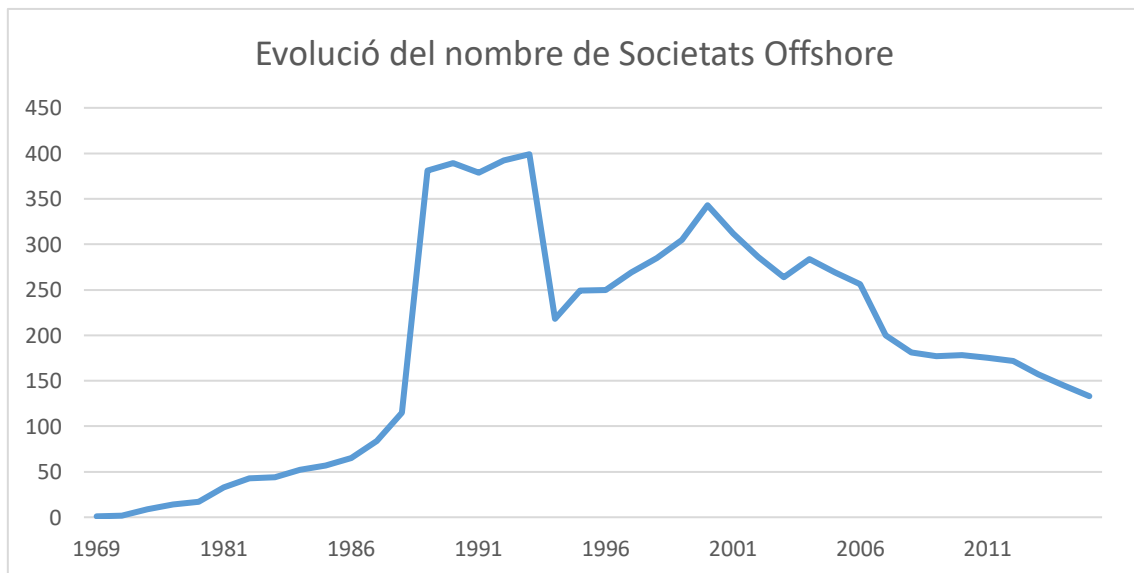
Com es pot observar en tots dos gràfics, tot i que l'ordre no coincideix, els mateixos paradisos fiscals són els que tenen major nombre de societats registrades, sis en el cas

de les societats espanyoles, amb un empat en el cinquè i sisè paradís amb major aflluència.

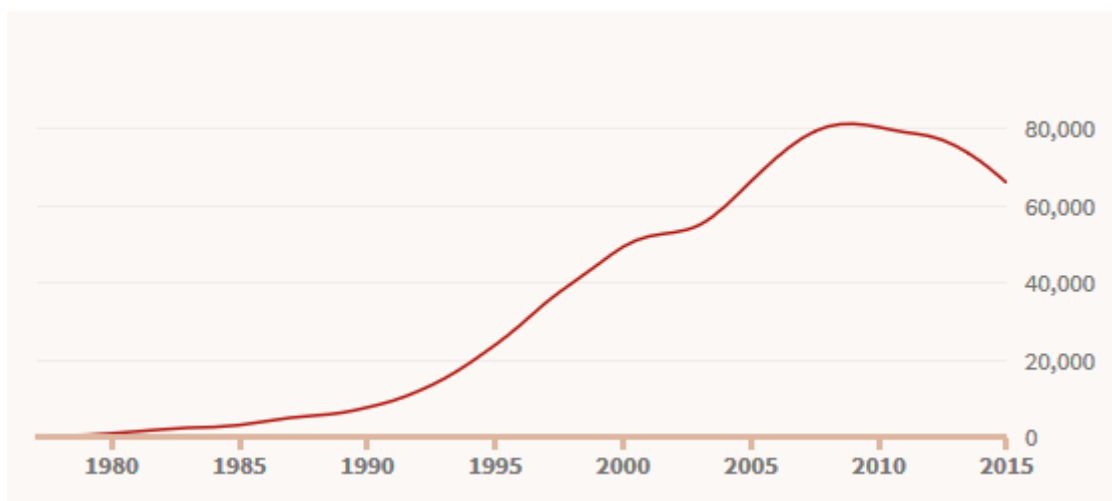
Destaca Panamà com a paradís fiscal més escollit per la creació de societats *offshore*, seguit per Niue i les *British Virgin Islands*. Aquest últim encara que no s'ha destacat molt en els gràfics i anàlisi anteriors, és el segon paradís fiscal que va començar a tenir societats *offshore* espanyoles i, tot i no registrar grans increments de societats actives, s'ha mantingut relativament constant al llarg de les dècades.

Pel que fa al comportament global, Panamà i *British Virgin Islands* disposen d'un registre molt més voluminós de societats *offshore* registrades. Al contrari que en el cas de les societats espanyoles, el paradís fiscal més utilitzat són les *British Virgin Islands* seguit de Panamà.

3.2.5 Evolució anual del total de les societats *offshore* actives presents als Papers de Panamà



Gràfic 10. Evolució de les societats offshore. Dades per a SPAIN. Font: Elaboració pròpia.



Gràfic 11. Evolució de les societats offshore. Dades globals. Font: Web dels Panama Papers.

Tot i que en els anteriors apartats s'han remarcat certes similituds entre el comportament de les societats *offshore* a nivell global i a nivell espanyol, en aquests gràfics que representen el comportament del total de societats *offshore* actives sense distingir per paradís fiscal, s'observa que no hi ha tanta similitud.

El pic de les societats panamenyes té molt pes en les societats espanyoles, mentre que en el gràfic global no es pot ni apreciar, de la mateixa manera passa amb el pic de les societats de Niue. Sí que sembla però que hi ha un moment d'estancament en el creixement de societats actives al 2001 que podria tenir relació amb la caiguda de les societats *offshore* creades a les Bahames.

On sí que hi ha una diferència total en el comportament de les societats actives, ja esmentat a l'apartat 3.3, és a partir del 2001. El nombre societats *offshore* espanyoles comença a decaure sense gairebé cap pujada fins a l'actualitat mentre que el comportament global té tendència a continuar augmentant fins al 2009 on comença a disminuir progressivament fins al 2013, on la tendència és molt més forta.

4 MODELITZACIÓ I COMPARACIÓ ANALÍTICA

En aquest apartat es realitzarà una comparació del comportament de les societats espanyoles i de la globalitat de les societats dels papers de Panamà mitjançant els models que millor s'ajusten a les dades. Com a l'anàlisi gràfica, la comparativa es realitzarà per als diferents conceptes clau, afegint una anàlisi per al temps mitjà de vida de les societats en funció de l'any de creació.

El procediment dut a terme és el mateix per a cada concepte clau:

1. Les dades s'han estandarditzat per tal de que estiguin a la mateixa escala i sigui possible dur a terme la comparació, ja que la població de les societats espanyoles és de l'ordre de milers, de centenars o inclús desenes si ho disgreguem per paradís fiscal, mentre que l'ordre de la globalitat de les dades és de centenars de milers o desenes de milers si ho disgreguem per paradís fiscal. S'han agafat només les dades globals corresponents al període de temps de les dades referents a les societats espanyoles.
2. S'ha realitzat la modelització de les dades trobant el model més complet, que millor expliqui i ajusti les dades. Per la comprovació d'aquest fet s'ha dut a terme el test F sota la hipòtesis nul·la.

$$F = \frac{MS_M}{MS_R} \quad \text{on} \quad MS_M = \frac{SS_M}{df_M}, MS_R = \frac{SS_R}{df_R}$$

Per tant, el model s'ajustarà millor i serà més explicatiu de les dades quan l'estadístic F sigui superior a 1, ja que voldrà dir que la millora de la predicció serà superior a la diferència entre el model i les dades explicades. També es prendrà com a referència el coeficient de determinació R^2 , que mesura el percentatge de variabilitat de les dades explicat pel model.

3. S'ha fet una comparació gràfica amb les dades i amb el model que millor ajusta superposat al gràfic de cada població respectivament.
4. Finalment, s'ha realitzat un test per comprovar si les variàncies dels models respecte les dades poden ser considerades iguals o no.

El test utilitzat per aquesta comprovació és el test F per a la comparació de variàncies aplicat al model lineal respecte les dades. Les hipòtesis del test són les següents:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

I l'estadístic del test és $F = S_1^2/S_2^2$

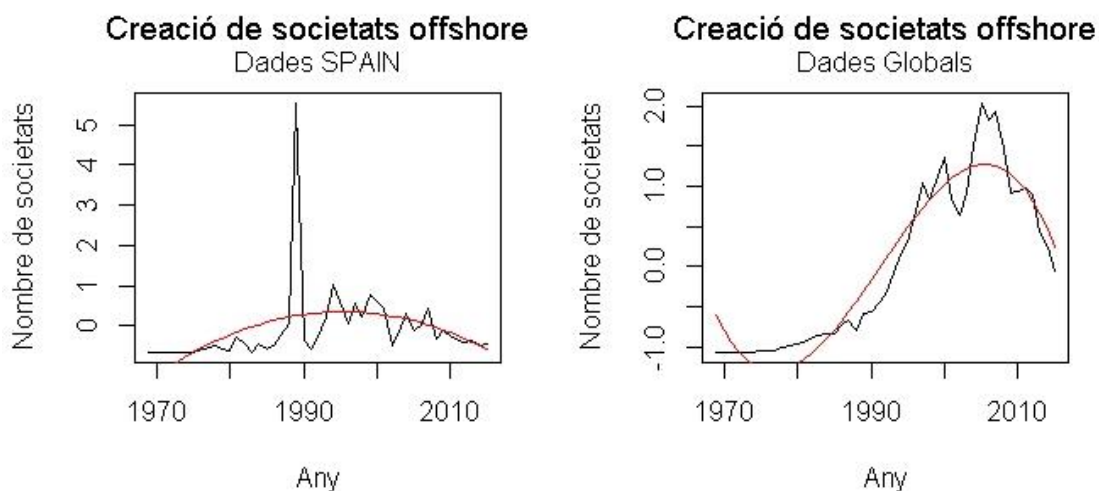
A continuació es procedeix a l'aplicació del procediment descrit pels diferents conceptes claus.

4.1 Evolució del nombre de societats *offshore* creades

A continuació s'analitzarà evolució del nombre de societats *offshore* creades amb dos casos diferents, el primer amb la totalitat de les dades de la societat espanyola i el segon eliminant l'outlier corresponent a l'any 1989.

4.1.1 Evolució del nombre de societats *offshore* creades

Al plot es pot comprovar directament com no hi ha un comportament semblant entre totes dues poblacions, l'outlier a la població de societats espanyoles desvirtua massa el gràfic, per tant gairebé podem concloure que les dues poblacions no es comporten igual respecte l'evolució anual del nombre de societats creades.



Gràfic 12. Nombre total de societats creades al llarg del temps. Dades per a SPAIN (esquerra) i globals (dreta). Font: Elaboració pròpia.

En el cas de les societats espanyoles, la modelització de les dades s'ha realitzat mitjançant un model polinòmic d'orde dos, ja que es tractava del model que millor ajustava les dades.

```
Call:
lm(formula = scale(count_spain) ~ poly(as.numeric(data_spain),
  2, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.8825 -0.3600 -0.1072  0.1315  5.2857

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.651e+03  4.160e+03  -2.320  0.0260 *
poly(as.numeric(data_spain), 2, raw = TRUE)1  9.673e+00  4.172e+00  2.319  0.0260 *
poly(as.numeric(data_spain), 2, raw = TRUE)2 -2.424e-03  1.046e-03  -2.317  0.0261 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9578 on 37 degrees of freedom
Multiple R-squared:  0.1297,    Adjusted R-squared:  0.08261
F-statistic: 2.756 on 2 and 37 DF,  p-value: 0.07661
```

Com es pot comprovar a l'output del model en R, tant el terme independent com els paràmetres són significatius sota una nivell confiança del 95%. El coeficient de determinació és de 0,08 i el p-valor resultant de l'estadístic F és de 0,07, fet que ens indica que, com es podia observar al plot, tot i ser el model que millor ajusta les dades, no és un bon model explicatiu de les mateixes.

En el cas de la població global, la modelització de les dades s'ha realitzat mitjançant un model polinòmic d'orde tres, que com es pot comprovar al plot, ajusta molt millor l'evolució anual del nombre de societats creades que no en el cas de la població espanyola.

```
Call:
lm(formula = scale(fonsf$count_fons) ~ poly(as.numeric(fonsf$data_fons),
  3, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.52578 -0.25466 -0.00307  0.22640  0.76083

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.721e+06  1.832e+05   9.392 5.59e-12 ***
poly(as.numeric(fonsf$data_fons), 3, raw = TRUE)1 -2.593e+03  2.759e+02  -9.397 5.51e-12 ***
poly(as.numeric(fonsf$data_fons), 3, raw = TRUE)2  1.302e+00  1.385e-01   9.401 5.44e-12 ***
poly(as.numeric(fonsf$data_fons), 3, raw = TRUE)3 -2.180e-04  2.318e-05  -9.405 5.38e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3108 on 43 degrees of freedom
Multiple R-squared:  0.9097,    Adjusted R-squared:  0.9034
F-statistic: 144.4 on 3 and 43 DF,  p-value: < 2.2e-16
```

El resultat obtingut del model, tant el terme independent com la resta de paràmetres, són significatius a tots els nivells de confiança. El coeficient de determinació, del 0,90, i

l'estadístic de contrast F de 144,4 amb 3 i 43 g.ll., amb un p-valor de pràcticament zero, ens indiquen que efectivament aquest model ajusta molt bé les dades i és un bon model explicatiu de les mateixes.

Finalment s'ha realitzat el test F descrit per comparar si les variàncies dels models són iguals.

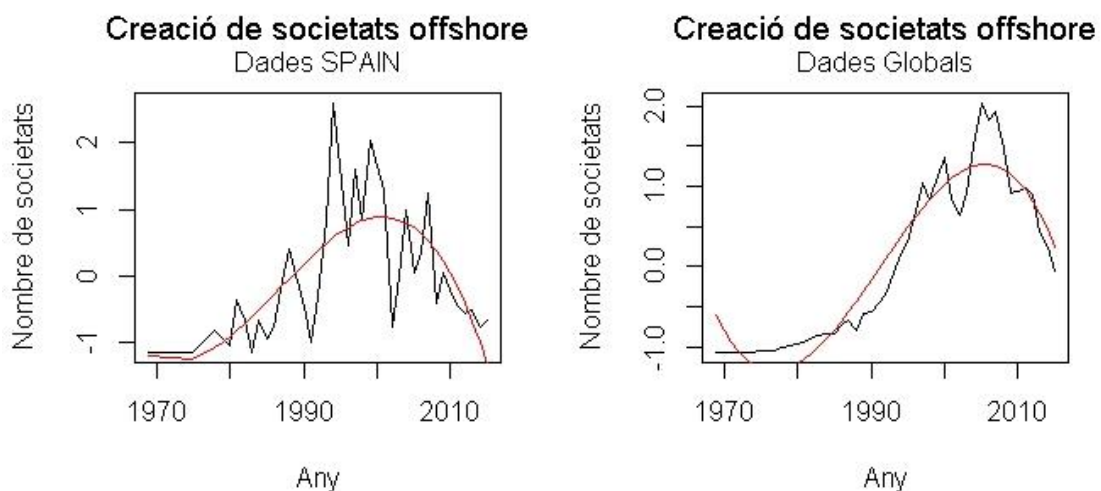
```

F test to compare two variances

data:  lm1 and lm2
F = 9.4957, num df = 37, denom df = 43, p-value = 3.401e-11
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 5.09261 17.97437
sample estimates:
ratio of variances
 9.495745
    
```

Amb un nivell de significació del 0,05 i com es pot esperar a la vista del plot i dels models emprats, l'estadístic de contrast F és 9,49 amb 37 i 43 g.ll., respectivament, i un p-valor pràcticament zero, acceptem la hipòtesi alternativa. Tenim suficients evidències per afirmar que les variàncies de tots dos models són diferents i per tant com a complement del plot, no es comporten de la mateixa manera.

4.1.2 Evolució del nombre de societat *offshore* creades sense l'outlier



Gràfic 13. Nombre total de societats creades al llarg del temps. Dades per a SPAIN sense l'outlier (esquerra) i globals (dreta). Font: Elaboració pròpia.

Degut a l'impacte de l'outlier sobre les dades i sobre el model utilitzat a l'apartat 4.1.1, s'ha optat per la seva eliminació per tal de poder repetir l'anàlisi i veure el comportament de les societats espanyoles sense la influència del mateix.

En aquest cas, sí que es pot observar al plot que totes dues poblacions tenen una tendència semblant i que tots dos models es comporten de la mateixa manera sense tenir en compte l'outlier a la població espanyola, tot i que representa prop d'un 20% de la població total.

El model utilitzat en aquest cas és d'ordre tres, igual que l'utilitzat a la modelització de la població global. Com es pot comprovar al plot, el model ajusta molt millor les dades que no el model utilitzat amb l'outlier a les dades.

```
Call:
lm(formula = scale(count_spain[-14]) ~ poly(as.numeric(data_spain[-14]),
      3, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-1.64790 -0.39885  0.00321  0.37872  2.01905

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.497e+06  4.705e+05   3.182  0.00306 **
poly(as.numeric(data_spain[-14]), 3, raw = TRUE)1 -2.261e+03  7.082e+02  -3.193  0.00297 **
poly(as.numeric(data_spain[-14]), 3, raw = TRUE)2  1.138e+00  3.554e-01   3.204  0.00289 **
poly(as.numeric(data_spain[-14]), 3, raw = TRUE)3 -1.910e-04  5.943e-05  -3.214  0.00281 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7132 on 35 degrees of freedom
Multiple R-squared:  0.5314,    Adjusted R-squared:  0.4913
F-statistic: 13.23 on 3 and 35 DF,  p-value: 6.225e-06
```

El resultat del model respecte les dades és que, tant el terme independent com la resta de paràmetres són significatius, i tant el coeficient de determinació, de 0,49, com l'estadístic F, de 13,23 amb 3 i 35 g.l., i un p-valor de $6,22e^{-6}$ indiquen que el model és molt millor que no l'emprat amb l'outlier a les dades. No és tan bo com el de la població global, però podem afirmar que ajusta bé les dades.

El model de la població global ja ha estat analitzat al subapartat 4.1.1.

F test to compare two variances

```
data:  lm1 and lm2
F = 5.2656, num df = 35, denom df = 43, p-value = 5.491e-07
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.804724 10.097819
sample estimates:
ratio of variances
 5.265616
```

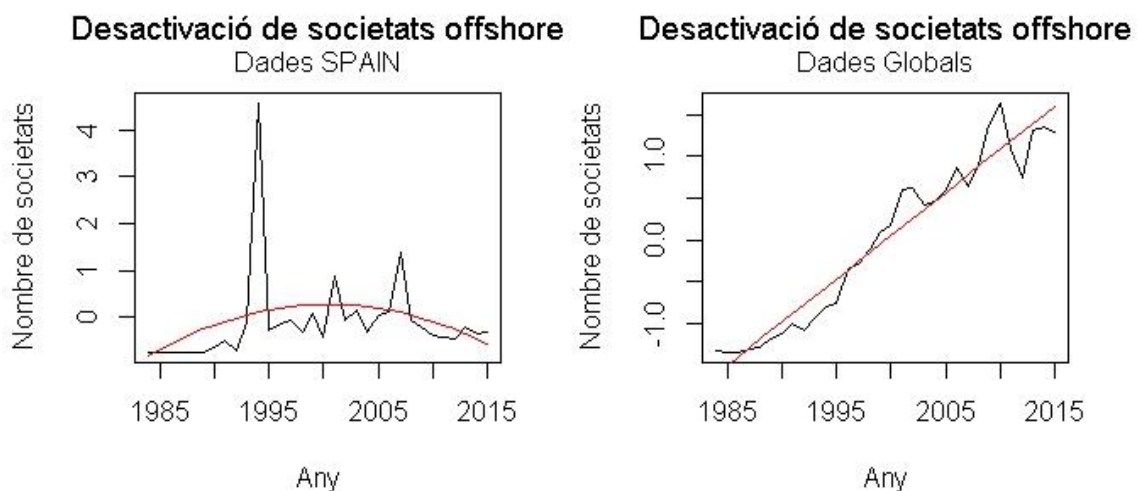

El resultat del test F per la comparació de les variàncies, tot i que la tendència de totes dues poblacions és la mateixa, amb un estadístic de contrast F igual a 5,26 amb 35 i 43 g.l., i un p-valor de $5,49e^{-7}$, indica que les variàncies de totes dues poblacions són diferents. Acceptem la hipòtesis alternativa.

Per tant es conclou que totes dues poblacions, tot i traient la dada influent, no es comporten de la mateixa manera. El nombre de noves societats espanyoles cau de manera constant a partir de 1994 amb un comportament força irregular d'un any a l'altre, mentre que el nombre de noves societats a nivell global segueix en continu creixement fins a 2005, on comença un decreixement molt més pronunciat i constant que en el cas de les societats espanyoles.

4.2 Evolució anual del nombre societats desactivades

En el segon subapartat s'analitzarà evolució del nombre de societats *offshore* desactivades, i de la mateixa manera que al punt 4.1 es realitzarà en dos casos diferents, el primer amb la totalitat de les dades de les societats espanyola i el segon eliminant l'outlier corresponent a l'any 1994.

4.2.1 Evolució anual del nombre societats desactivades



Gràfic 14. Nombre total de societats desactivades al llarg del temps. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

En aquest plot, de la mateixa manera que al subapartat 4.1.1, a primera vista es pot observar com no hi ha un comportament de les dues poblacions semblant. Igual que en el subpunt esmentat, la influència de l'outlier desvirtua el gràfic i per tant desvirtua el model. Podríem concloure que les dues poblacions tenen comportaments no iguals.

En el cas de les societats espanyoles, cap model polinòmic ha estat satisfactori (en termes d'ajust global) ni ha permès modelitzar les dades, degut en gran part a la influència de l'outlier. El millor model, tot i així, aconseguit ha estat quadràtic, amb cap paràmetre significatiu i un coeficient de determinació pràcticament zero.

```
Call:
lm(formula = scale(count_spain2) ~ poly(as.numeric(data_spain2),
    2, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.6969 -0.4109 -0.1922 -0.0064  4.4434

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.597e+04  1.085e+04  -1.472   0.153
poly(as.numeric(data_spain2), 2, raw = TRUE)1  1.596e+01  1.084e+01   1.472   0.153
poly(as.numeric(data_spain2), 2, raw = TRUE)2  -3.990e-03  2.709e-03  -1.473   0.153

Residual standard error: 0.9966 on 25 degrees of freedom
Multiple R-squared:  0.08042, Adjusted R-squared:  0.006851
F-statistic: 1.093 on 2 and 25 DF, p-value: 0.3507
```

El model utilitzat en el cas de la població global ha estat un model lineal que s'ajusta bé a les dades, en continu creixement.

```
Call:
lm(formula = scale(fonsf2$count_fons) ~ poly(as.numeric(fonsf2$data_fons),
    1, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.54279 -0.12978 -0.00371  0.13577  0.54921

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.070e+02  9.300e+00  -22.25  <2e-16 ***
poly(as.numeric(fonsf2$data_fons), 1, raw = TRUE)  1.035e-01  4.651e-03   22.25  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

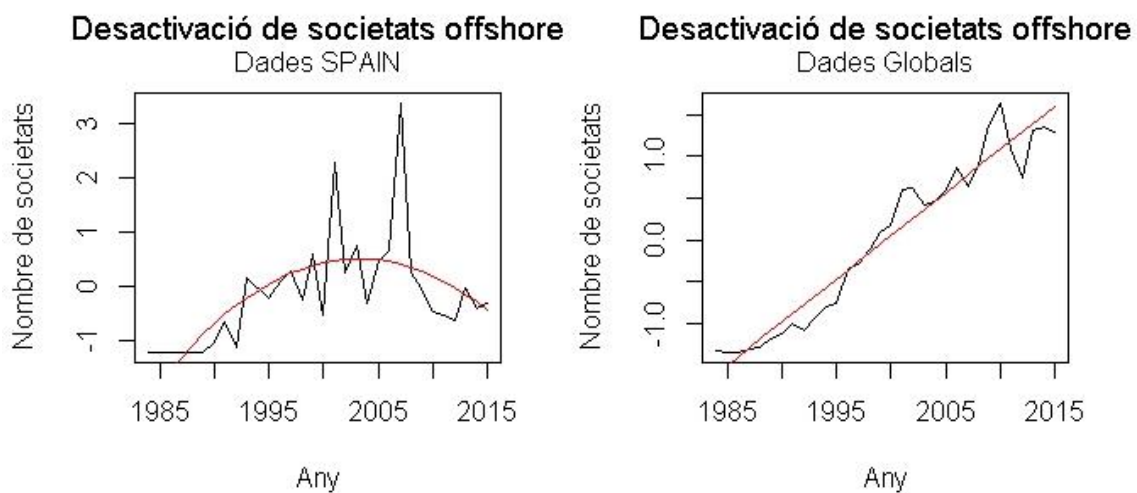
Residual standard error: 0.2429 on 30 degrees of freedom
Multiple R-squared:  0.9429, Adjusted R-squared:  0.941
F-statistic: 495.3 on 1 and 30 DF, p-value: < 2.2e-16
```

El model té tots els paràmetres significatius a tot nivell de significació, amb un coeficient de determinació de 0,94 i un p-valor de zero pel que fa l'estadístic F. Pel que podem afirmar que és un model que ajusta molt bé les dades i amb molt bona capacitat explicativa.

En base a que la modelització de les societats espanyoles desactivades no ha estat possible, no es procedeix al test per comprovar si les variàncies de tot dos models són iguals i es conclou que totes dues poblacions no tenen el mateix comportament.

4.2.2 Evolució anual del nombre societats desactivades sense l'outlier

Degut a la influència de l'outlier sobre les dades i sobre el model utilitzat al subapartat 4.2.1, s'ha dut a terme la seva eliminació de les dades per repetir l'anàlisi i veure el comportament de les societats espanyoles sense la influència d'aquest.



Gràfic 15. Nombre total de societats desactivades al llarg del temps. Dades per a SPAIN (sense l'outlier) i globals (dreta). Font: Elaboració pròpia.

Com a les dades de les societats espanyoles creades, l'outlier eliminat representa prop del 20% de la població. En aquest cas tot i haver-lo eliminat es pot observar al plot que no hi ha una tendència ni un comportament semblant i tots dos models es comporten de diferent manera, per tant sense tenir en compte l'outlier a la població espanyola, no sembla que hi pugui haver igualtat entre totes dues poblacions.

La modelització de les dades sense l'outlier s'ha realitzat amb un model d'ordre 2, on tots els seus paràmetres són significatius inclòs l'estadístic F de 5,89 amb 2 i 24 g.l., aconseguint un coeficient de determinació de 0,27, no gaire elevat per tant no es pot dir que sigui un bon model que ajusti bé les dades.

```

Call:
lm(formula = scale(count_spain2[-7]) ~ poly(as.numeric(data_spain2[-7]),
      2, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.9728 -0.4617 -0.1077  0.1734  2.9822

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.777e+04  9.295e+03  -2.987  0.00639 **
poly(as.numeric(data_spain2[-7]), 2, raw = TRUE)1  2.772e+01  9.291e+00   2.984  0.00645 **
poly(as.numeric(data_spain2[-7]), 2, raw = TRUE)2 -6.920e-03  2.322e-03  -2.980  0.00650 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8523 on 24 degrees of freedom
Multiple R-squared:  0.3295,    Adjusted R-squared:  0.2736
F-statistic: 5.896 on 2 and 24 DF,  p-value: 0.00826

```

El model de la població global ja ha estat analitzat al subapartat 4.2.1.

Finalment, el test F per la comparació de les variàncies ha donat com ha resultat un estadístic de contrast F igual a 12,90 amb 24 i 30 g.ll., un p-valor de $1,69e^{-9}$. Per tant acceptem la hipòtesis alternativa, les variàncies de totes dues poblacions són diferents.

```

F test to compare two variances

data:  lm1 and lm2
F = 12.309, num df = 24, denom df = 30, p-value = 1.692e-09
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 5.762879 27.189862
sample estimates:
ratio of variances
 12.30881

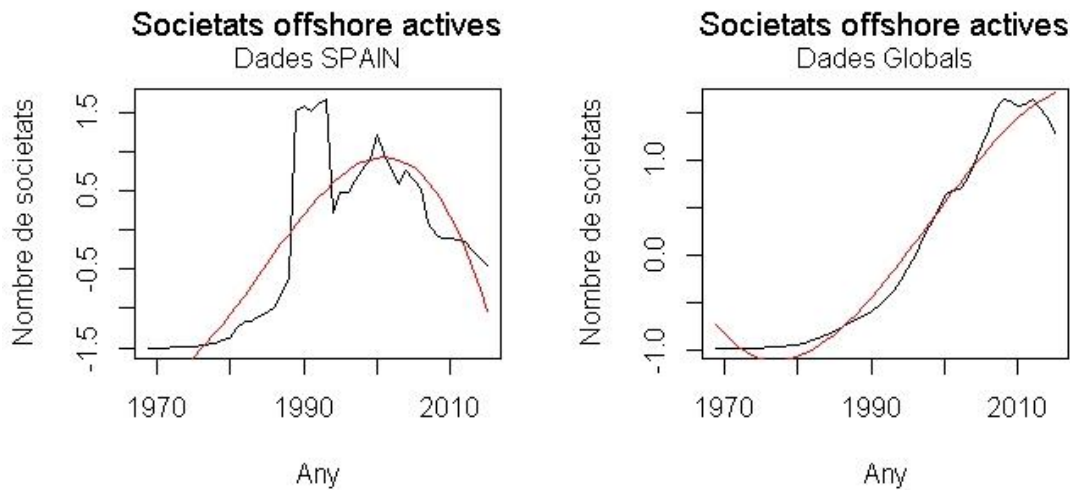
```

Es conclou en aquest apartat que totes dues poblacions, tot i traient la dada influent, no es comporten de la mateixa manera. El nombre de societats espanyoles desactivades té dos pics de societats desactivades a 2001 i 2007. Amb un comportament força irregular d'un any a l'altre, i a mesura que augmentem el moment en el temps (es creen menys societats) cau el nombre de societats desactivades, mentre que el nombre de societats desactivades a nivell global té un creixement constant en la totalitat de les dades.

4.3 Societats actives per any

Es pot observar en el plot com totes dues poblacions tenen una tendència ascendent durant la major part del temps reflectit al plot i que la població de societats espanyoles comença a decreixer anys abans que la població global. Destaca el pic de gran

quantitat de societats actives a la població espanyola cap al 1990. Pel què fa als models, a priori tots dos s'ajusten bé a les poblacions respectives.



Gràfic 16. Nombre total de societats actives al llarg del temps. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

La modelització de les societats espanyoles actives s'ha realitzat a partir d'un model d'ordre 3, que millor ajustava els canvis de tendència de les dades.

```
Call:
lm(formula = scale(spainf4$x) ~ poly(as.numeric(spainf4$acti),
  3, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.85683 -0.08721 -0.00126  0.13449  0.94442

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.026e+06  1.834e+05   5.593 2.43e-06 ***
poly(as.numeric(spainf4$acti), 3, raw = TRUE)1 -1.538e+03  2.761e+02  -5.572 2.60e-06 ***
poly(as.numeric(spainf4$acti), 3, raw = TRUE)2  7.688e-01  1.385e-01   5.550 2.77e-06 ***
poly(as.numeric(spainf4$acti), 3, raw = TRUE)3 -1.281e-04  2.317e-05  -5.529 2.96e-06 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2792 on 36 degrees of freedom
Multiple R-squared:  0.9281,    Adjusted R-squared:  0.9221
F-statistic: 154.8 on 3 and 36 DF,  p-value: < 2.2e-16
```

S'ha obtingut un model amb tots els paràmetres significatius, un coeficient de determinació de 0,92 i un p-valor de pràcticament zero resultant de l'estadístic F igual a 154,8 amb 3 i 36 g.l., per tant es pot afirmar que és un molt bon model, que ajusta bé les dades.

Pel que fa al model de la població global s'ha ajustat amb un model d'ordre 3 també, que degut a la tendència de les dades, es pot comprovar al plot que les ajusta bé.

```

Call:
lm(formula = scale(fonsf3$acti) ~ poly(as.numeric(fonsf3$data_fons),
3, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.41383 -0.08867  0.00739  0.11219  0.34105

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.439e+05  9.370e+04   6.872 1.97e-08 ***
poly(as.numeric(fonsf3$data_fons), 3, raw = TRUE)1 -9.671e+02  1.411e+02  -6.853 2.10e-08 ***
poly(as.numeric(fonsf3$data_fons), 3, raw = TRUE)2  4.841e-01  7.085e-02   6.834 2.24e-08 ***
poly(as.numeric(fonsf3$data_fons), 3, raw = TRUE)3 -8.078e-05  1.186e-05  -6.814 2.40e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.159 on 43 degrees of freedom
Multiple R-squared:  0.9764,    Adjusted R-squared:  0.9747
F-statistic: 592.5 on 3 and 43 DF,  p-value: < 2.2e-16

```

Com s'esperava, el resultat del model té tots els paràmetres significatius a qualsevol nivell de significació, amb un coeficient de determinació del 0,97, un estadístic de contrast de 592,5 amb 3 i 43 g.ll. i un p-valor de pràcticament zero, podem afirmar que és un bon model explicatiu de les dades.

Finalment, s'ha realitzat el test F per a la comparació de variàncies del dos models per comprovar si aquestes es comporten de la mateixa manera.

```

      F test to compare two variances

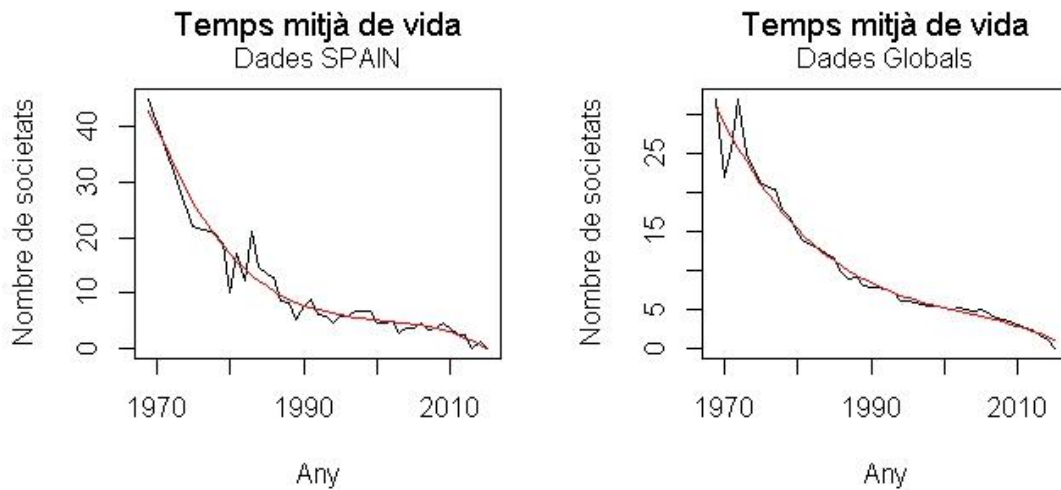
data:  lm1 and lm2
F = 14.289, num df = 36, denom df = 43, p-value = 2.487e-14
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 7.637893 27.220286
sample estimates:
ratio of variances
 14.28937

```

Com a resultat del test s'obté un p-valor pràcticament nul, amb un estadístic de contrast F de 14,28 amb 36 i 43 g.ll. Per tant, tenim suficients evidències per rebutjar la hipòtesis nul·la, les variàncies de les dues poblacions són diferents. Concloem que les dues poblacions per a l'evolució del nombre de societats actives per any no tenen un mateix comportament. Tot i tenir una clara tendència creixent, per a la població de societats espanyoles hi ha un canvi de tendència molt més pronunciat i molt més d'hora que no a la població global de societats actives.

4.4 Temps mitjà de vida de les societats

El plot del temps mitjà de vida de les societats amb data de desactivació, sí que mostra un comportament pràcticament igual entre totes dues poblacions pel que fa a la tendència de les dades, amb uns models que tenen un comportament també molt semblant.



Gràfic 17. Temps mitjà de vida al llarg del temps. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

El model que millor ajusta el temps mitjà de vida de les societats espanyoles és un model d'ordre 3 i que com s'ha pogut comprovar en el plot ajusta molt bé a les dades.

```
Call:
lm(formula = scale(spainf4$x) ~ poly(as.numeric(spainf4$acti),
  3, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.85683 -0.08721 -0.00126  0.13449  0.94442

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.026e+06  1.834e+05   5.593 2.43e-06 ***
poly(as.numeric(spainf4$acti), 3, raw = TRUE)1 -1.538e+03  2.761e+02  -5.572 2.60e-06 ***
poly(as.numeric(spainf4$acti), 3, raw = TRUE)2  7.688e-01  1.385e-01   5.550 2.77e-06 ***
poly(as.numeric(spainf4$acti), 3, raw = TRUE)3 -1.281e-04  2.317e-05  -5.529 2.96e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2792 on 36 degrees of freedom
Multiple R-squared:  0.9281,    Adjusted R-squared:  0.9221
F-statistic: 154.8 on 3 and 36 DF,  p-value: < 2.2e-16
```

Com es pot comprovar tots els paràmetres son significatius a tots els nivells de significació, així com l'estadístic F de 154,8 amb 3 i 36 g.l. i un p-valor de zero ens indiquen que efectivament és un model que s'ajusta molt bé a les dades, el coeficient

de determinació, de 0,92, ens mostra també l'efectivitat del model i que és un bon model explicatiu de les dades.

Com en el cas de les societats espanyoles, el model que millor ajusta el temps mitjà de vida de les societats globals és un model d'ordre 3 i que, com en el primer cas, ajusta molt bé a les dades.

```
Call:
lm(formula = scale(fonsf4$x) ~ poly(as.numeric(fonsf4$acti),
  3, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.84973 -0.05140  0.00783  0.04762  0.76628

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.716e+05  1.134e+05   4.158 0.000150 ***
poly(as.numeric(fonsf4$acti), 3, raw = TRUE)1 -7.063e+02  1.708e+02  -4.135 0.000162 ***
poly(as.numeric(fonsf4$acti), 3, raw = TRUE)2  3.526e-01  8.575e-02   4.112 0.000174 ***
poly(as.numeric(fonsf4$acti), 3, raw = TRUE)3 -5.867e-05  1.435e-05  -4.089 0.000186 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1924 on 43 degrees of freedom
Multiple R-squared:  0.9654,    Adjusted R-squared:  0.963
F-statistic: 399.9 on 3 and 43 DF,  p-value: < 2.2e-16
```

De la mateixa manera que en la població de les societats espanyoles, tots els paràmetres són significatius a tots els nivells de significació, l'estadístic F és de 399,9 amb 3 i 43 g.l. i el p-valor de zero indiquen que és un model que s'ajusta molt bé a les dades. El coeficient de determinació, de 0,96, ens indica també l'efectivitat del model i que és un bon model explicatiu de les dades.

Com a última comprovació per veure si les dues poblacions tenen una mateixa variància i per tant concloure si es comporten igual, es realitza el test F per comparació de variàncies.

```
F test to compare two variances

data:  lm1 and lm2
F = 2.1052, num df = 36, denom df = 43, p-value = 0.02013
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.125269 4.010287
sample estimates:
ratio of variances
 2.105213
```

El resultat del test dona com a resultat un p-valor de 0,02 amb un estadístic de contrast F de 2,10 amb 36 i 43 g.l, respectivament. Per tant, a un nivell de significació del 0,05, rebutgem la hipòtesis nul·la, tenim prou evidències per indicar que totes dues

poblacions no es comporten de manera semblant. Tot i que les dues poblacions sí que tenen una tendència pràcticament igual, per a la població de les societats espanyoles, el comportament de les dades és més irregular que no el de la població global, d'aquí la diferència significativa entre el comportament de totes dues poblacions.

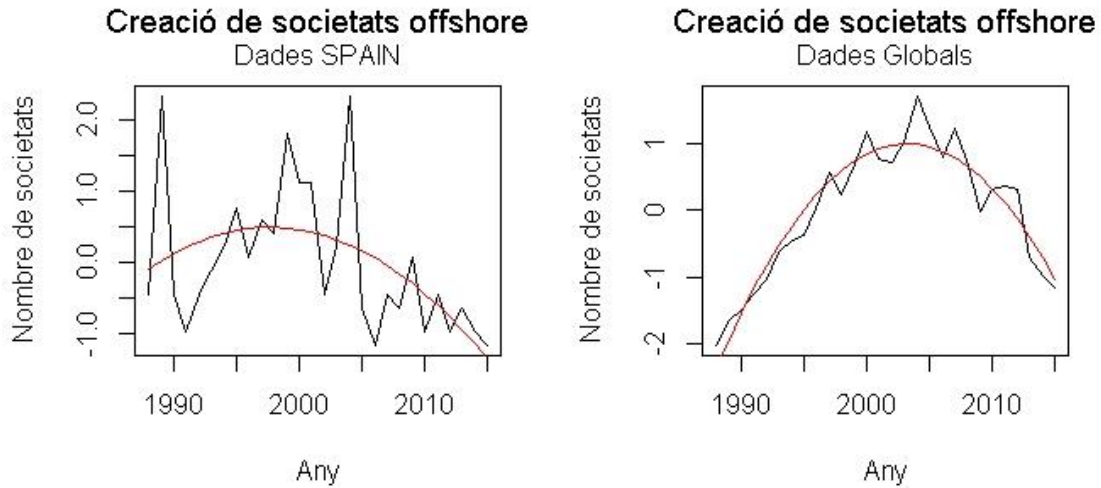
Cal destacar que si el contrast es fes per un nivell de significació de 0,01, no podríem rebutjar la hipòtesis nul·la, no podríem dir que les poblacions tenen una variància diferent, i per tant, dir que tenen un comportament igual.

4.5 Comparació de l'evolució anual de societats creades per paradís fiscal

En aquest apartat, a partir del gràfic on apareix l'evolució anual de les societats creades dividit per paradís fiscal, s'ha procedit a fer una anàlisi per separat del comportament de cada paradís fiscal. Per poder dur a terme aquesta anàlisi només s'han tingut en compte aquells paradisos fiscals que formen part de totes dues poblacions, i com a segona condició, només d'aquells paradisos fiscals amb suficients dades per poder generar un model representatiu de les mateixes. Per tant, només s'analitzaran els paradisos fiscals de les *British Virgin Island* i de Niue.

Per poder procedir a la comparació per paradís fiscal, es va crear una funció d'R que permetia realitzar els models i les validacions dels mateixos, així com el resultat del test F per la comparació de les variàncies de totes dues poblacions, de manera que a l'explicació de cada model s'indicarà el tipus de model utilitzat tot i que no es vegi en l'output d'R.

4.5.1 Paradís fiscal 1: *British Virgin Islands*



Gràfic 18. Nombre de societats creades al paradís fiscal British Virgin Islands. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

Com es pot comprovar al plot, no hi ha un comportament semblant entre totes dues poblacions, gairebé podríem concloure que les dues poblacions no es comporten igual respecte l'evolució anual del nombre de societats creades.

El model utilitzat per a la població de societats espanyoles és un model d'ordre 2 que com es pot comprovar al plot, no ajusta bé les dades degut a la irregularitat de les mateixes.

```
Call:
lm(formula = scale(pfsp$count) ~ poly(as.numeric(pfsp$data),
  n, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-1.2277 -0.4906 -0.1261  0.3150  2.3093

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.463e+04  1.143e+04  -2.155  0.0410 *
poly(as.numeric(pfsp$data), n, raw = TRUE)1  2.466e+01  1.142e+01   2.159  0.0406 *
poly(as.numeric(pfsp$data), n, raw = TRUE)2  -6.172e-03  2.853e-03  -2.164  0.0403 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8793 on 25 degrees of freedom
Multiple R-squared:  0.2842,    Adjusted R-squared:  0.2269
F-statistic: 4.962 on 2 and 25 DF,  p-value: 0.01532
```

Tots els paràmetres del model són significatius per a un nivell de significació del 0,05. El coeficient de determinació, de 0,22, i l'estadístic F igual a 4,96 amb 2 i 25 g.l., i un p-

valor de 0,015 ens permeten afirmar l'esperat al veure el plot, que el model no ajusta bé les dades.

El model utilitzat per a la població global, es tracta també d'un model d'ordre 2 que com es pot comprovar al plot, té un ajust molt bo de les dades.

```
Call:
lm(formula = scale(pffs$count) ~ poly(as.numeric(pffs$data),
  n2, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.53311 -0.20858 -0.07033  0.26825  0.71817

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.838e+04  4.048e+03  -14.42 1.27e-13 ***
poly(as.numeric(pffs$data), n2, raw = TRUE)1  5.829e+01  4.045e+00   14.41 1.29e-13 ***
poly(as.numeric(pffs$data), n2, raw = TRUE)2 -1.455e-02  1.010e-03  -14.40 1.31e-13 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3114 on 25 degrees of freedom
Multiple R-squared:  0.9102,    Adjusted R-squared:  0.903
F-statistic: 126.7 on 2 and 25 DF,  p-value: 8.244e-14
```

Els paràmetres del model són significatius a qualsevol nivell de significació, el coeficient de determinació es de 0,90, l'estadístic F igual a 126,7 amb 2 i 25 g.ll., i un p-valor de pràcticament zero, podem afirmar que el model ajusta molt bé les dades.

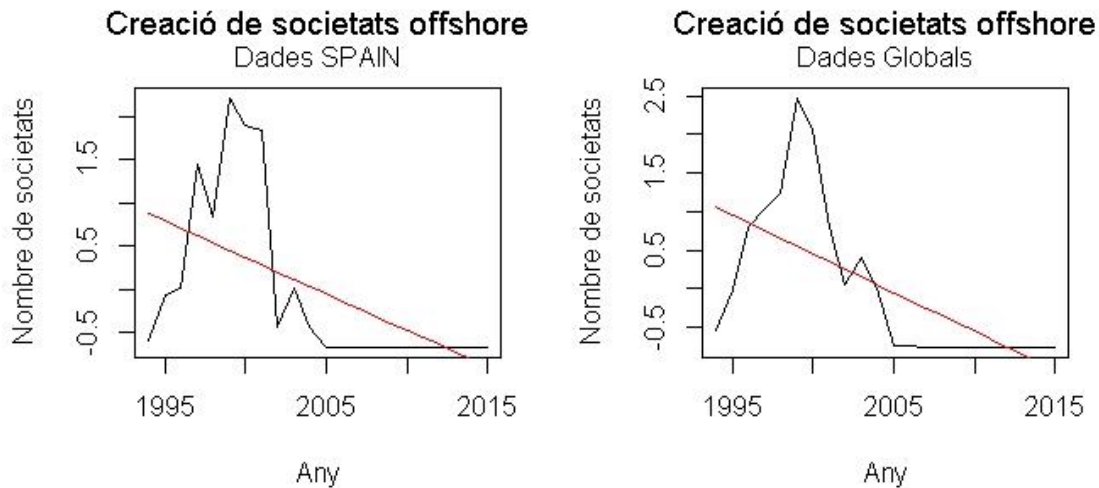
F test to compare two variances

```
data:  lm1 and lm2
F = 7.9711, num df = 25, denom df = 25, p-value = 1.87e-06
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 3.574004 17.777979
sample estimates:
ratio of variances
 7.971108
```

El resultat del test F per a la comparació de variàncies dóna un p-valor proper a zero, amb un estadístic de contrast F de 7,97 amb 25 i 25 g.ll. respectivament. Rebutgem la hipòtesis nul·la, les variàncies de totes dues poblacions són diferents.

Concloem per tant que totes dues poblacions es comporten de manera diferent.

4.5.2 Paradís fiscal 2: Niue



Gràfic 19. Nombre de societats creades al paradís fiscal Niue. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

El comportament de totes dues poblacions i de tots dos models és semblant pel que fa al comportament de les societats creades, com es pot comprovar al plot.

El model utilitzat per la població de societats espanyoles és un model d'ordre 1.

```
Call:
lm(formula = scale(pfsp$count) ~ poly(as.numeric(pfsp$data),
  n, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-1.4702 -0.5314 -0.1087  0.2295  1.7614

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  169.47371   57.74264   2.935  0.0085 **
poly(as.numeric(pfsp$data), n, raw = TRUE) -0.08455   0.02881  -2.935  0.0085 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.851 on 19 degrees of freedom
Multiple R-squared:  0.3119,    Adjusted R-squared:  0.2757
F-statistic: 8.614 on 1 and 19 DF,  p-value: 0.008498
```

Tant el terme independent com la resta de paràmetres són significatius per a un nivell de significació de 0,05. L'estadístic de contrast F igual a 8,61 amb 1 i 19 g.l., respectivament. Amb un coeficient de determinació de 0,27, indica que el model explica el 27% de la variabilitat de les dades. No podem dir que sigui un bon model.

El model utilitzat per la població global es tracta també d'un model d'ordre 2.

```

Call:
lm(formula = scale(pffs$count) ~ poly(as.numeric(pffs$data),
    n2, raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-1.60166 -0.38154 -0.05562  0.27896  1.91519

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      201.68505    52.25548     3.86 0.000977 ***
poly(as.numeric(pffs$data), n2, raw = TRUE)  -0.10062     0.02607    -3.86 0.000977 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7757 on 20 degrees of freedom
Multiple R-squared:  0.4269,    Adjusted R-squared:  0.3982
F-statistic: 14.9 on 1 and 20 DF,  p-value: 0.0009767

```

Tant el terme independent com la resta de paràmetres són significatius per a tots els nivells de significació, obtenint un coeficient de determinació de 0,39 així com l'estadístic F igual a 14,9 amb 1 i 20 g.ll., respectivament, i amb un p-valor de 0,0009. Com en el cas anterior, no podem dir que sigui un bon model.

Com es podia preveure al plot, el resultat del test F per a la comparació de les variàncies de totes dues poblacions, ha donat com a resultat un p-valor de 0,68 i un estadístic de contrast de 1,20 amb 19 i 20 g.ll., respectivament. Per tant tenim prou evidències per a acceptar la hipòtesis nul·la, les variàncies de totes dues poblacions són iguals.

```

F test to compare two variances

data:  lm1 and lm2
F = 1.2035, num df = 19, denom df = 20, p-value = 0.6835
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4848937 3.0196191
sample estimates:
ratio of variances
      1.203543

```

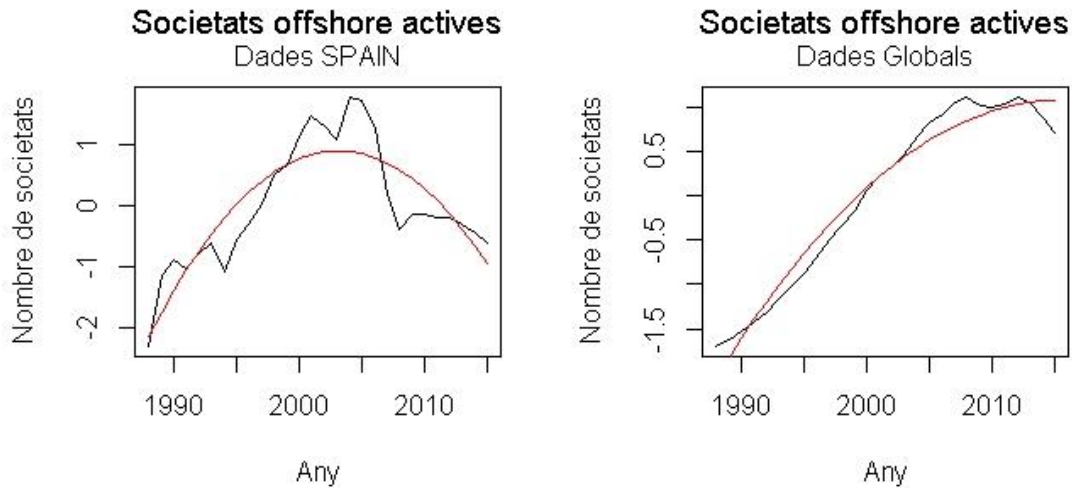
En base al test d'hipòtesis i al plot, concloem que totes dues poblacions es comporten de la mateixa manera al paradís fiscal Niue, pel que fa al nombre de societats creades.

4.6 Comparació de l'evolució anual de societats actives per paradís fiscal

Per tal de fer una comparativa més completa, a més de la comparativa per l'evolució anual del nombre de societats creades, s'ha realitzat també una comparativa per l'evolució anual del nombre de societats actives. Seguint les mateixes pautes i

requeriments que a l'apartat 4.5, s'ha procedit a l'anàlisi de l'evolució anual de societats actives als paradisos fiscals de les *British Virgin Islands*, Niue i a més Panamà.

4.6.1 Paradís fiscal 1: *British Virgin Islands*



Gràfic 20. Nombre de societats actives al paradís fiscal *British Virgin Islands*. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

Com es pot comprovar al plot, no hi ha un comportament semblant entre totes dues poblacions, tot i que totes dues tinguin una tendència ascendent amb una posterior tendència descendent, gairebé podríem concloure que les dues poblacions no es comporten igual respecte l'evolució anual del nombre de societats actives.

El model utilitzat per a la població de societats espanyoles és un model d'ordre 2 que com es pot comprovar al plot, ajusta bé les tendències de les dades.

```

Call:
lm(formula = scale(pfsp$acti) ~ poly(as.numeric(pfsp$data), n,
    raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.97086 -0.38469 -0.02122  0.38634  0.89276

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -5.327e+04  6.750e+03  -7.893 3.00e-08
poly(as.numeric(pfsp$data), n, raw = TRUE)1  5.319e+01  6.745e+00   7.886 3.05e-08
poly(as.numeric(pfsp$data), n, raw = TRUE)2 -1.328e-02  1.685e-03  -7.880 3.09e-08

(Intercept)                ***
poly(as.numeric(pfsp$data), n, raw = TRUE)1 ***
poly(as.numeric(pfsp$data), n, raw = TRUE)2 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5193 on 25 degrees of freedom
Multiple R-squared:  0.7503,    Adjusted R-squared:  0.7303
F-statistic: 37.55 on 2 and 25 DF,  p-value: 2.941e-08

```

Tots els paràmetres del model són significatius a qualsevol nivell de significació, i tant el coeficient de determinació, de 0,73, com l'estadístic F igual a 37,55 amb 2 i 25 g.ll., i un p-valor de $2,94e^{-8}$ ens permeten afirmar que el model ajusta bé les dades.

El model utilitzat per la població global es tracta també d'un model d'ordre 2 que com es pot comprovar al plot, té un ajust molt bo de les dades.

```

Call:
lm(formula = scale(pffs$acti) ~ poly(as.numeric(pffs$data), n2,
    raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.37438 -0.15412 -0.00401  0.11974  0.35835

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.703e+04  2.395e+03  -7.112 1.87e-07
poly(as.numeric(pffs$data), n2, raw = TRUE)1  1.690e+01  2.393e+00   7.063 2.10e-07
poly(as.numeric(pffs$data), n2, raw = TRUE)2 -4.194e-03  5.979e-04  -7.015 2.36e-07

(Intercept)                ***
poly(as.numeric(pffs$data), n2, raw = TRUE)1 ***
poly(as.numeric(pffs$data), n2, raw = TRUE)2 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1843 on 25 degrees of freedom
Multiple R-squared:  0.9686,    Adjusted R-squared:  0.966
F-statistic: 385 on 2 and 25 DF,  p-value: < 2.2e-16

```

Com en el model anterior, tots els paràmetres del model són significatius a qualsevol nivell de significació, el coeficient de determinació es de 0,96, l'estadístic F igual a 385 amb 2 i 25 g.ll., i un p-valor de pràcticament zero, podem afirmar que el model ajusta molt bé les dades.

El resultat del test F per la comparació de variàncies dóna un p-valor inferior a 0,01, amb un estadístic de contrast F de 7,94 amb 25 i 25 g.l., respectivament. Rebutgem la hipòtesi nul·la, les variàncies de totes dues poblacions són diferents.

```

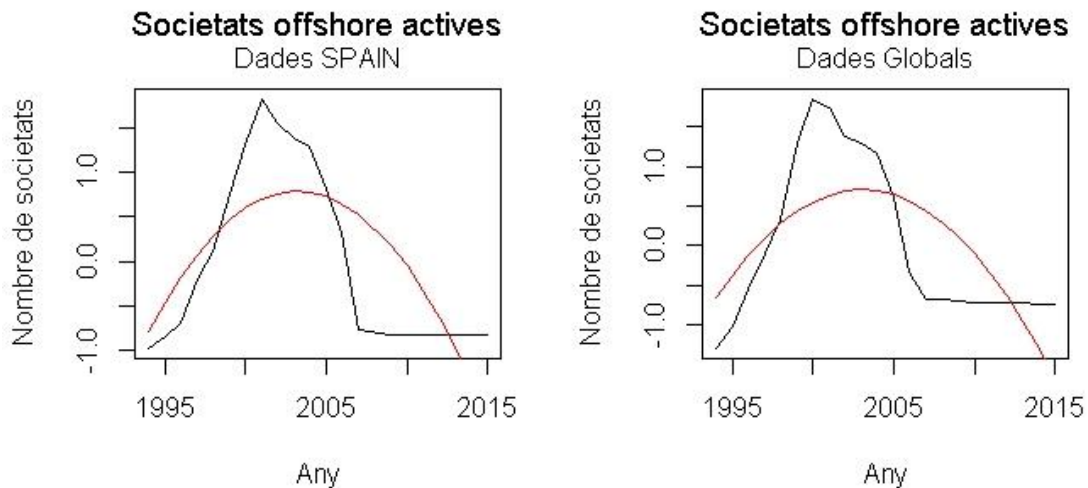
F test to compare two variances

data: lm1 and lm2
F = 7.9423, num df = 25, denom df = 25, p-value = 1.938e-06
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 3.561068 17.713633
sample estimates:
ratio of variances
 7.942257

```

Concloem per tant, que totes dues poblacions es comporten de manera diferent.

4.6.2 Paradís fiscal 2: Niue



Gràfic 21. Nombre de societats actives al paradís fiscal Niue. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

El comportament de totes dues poblacions és pràcticament igual pel que fa al comportament de les societats actives com es pot comprovar al plot. Tots dos models tenen el mateix comportament i són del mateix ordre.

El model utilitzat per la població de societats espanyoles és un model d'ordre 2.

```
Call:
lm(formula = scale(pfsp$acti) ~ poly(as.numeric(pfsp$data), n,
  raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-1.3061 -0.3903 -0.1850  0.5166  1.1230

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -7.400e+04  1.670e+04  -4.430  0.000323 ***
poly(as.numeric(pfsp$data), n, raw = TRUE)1  7.388e+01  1.667e+01   4.433  0.000321 ***
poly(as.numeric(pfsp$data), n, raw = TRUE)2  -1.844e-02  4.157e-03  -4.436  0.000319 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6894 on 18 degrees of freedom
Multiple R-squared:  0.5722,    Adjusted R-squared:  0.5247
F-statistic: 12.04 on 2 and 18 DF,  p-value: 0.0004798
```

Tant el terme independent com la resta de paràmetres són significatius per a tots els nivells de significació, així com l'estadístic F igual a 12,04 amb 2 i 18 g.ll., respectivament, amb un coeficient de determinació de 0,52. Indicant que el model explica el 52% de la variabilitat de les dades, no podem dir que és un model dolent.

El model utilitzat per la població global és també un model d'ordre 2.

```
Call:
lm(formula = scale(pffs$acti) ~ poly(as.numeric(pffs$data), n2,
  raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-1.11359 -0.62112 -0.06588  0.57590  1.28869

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -6.681e+04  1.821e+04  -3.668  0.00163 **
poly(as.numeric(pffs$data), n2, raw = TRUE)1  6.671e+01  1.817e+01   3.671  0.00162 **
poly(as.numeric(pffs$data), n2, raw = TRUE)2  -1.665e-02  4.533e-03  -3.673  0.00161 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.763 on 19 degrees of freedom
Multiple R-squared:  0.4732,    Adjusted R-squared:  0.4178
F-statistic: 8.535 on 2 and 19 DF,  p-value: 0.002267
```

Com al primer model, tant el terme independent com la resta de paràmetres són significatius per a tots els nivells de significació, obtenint un coeficient de determinació de 0,41 així com l'estadístic F igual a 8,53 amb 2 i 19 g.ll., respectivament. Amb un p-valor de 0,002, com en el cas anterior, no podem dir que sigui un model dolent.

El resultat del test F per a la comparació de les variàncies de totes dues poblacions ha donat com a resultat un p-valor de 0,67 i un estadístic de contrast de 0,81 amb 18 i 19

g.II., respectivament. Per tant tenim prou evidències per a acceptar la hipòtesis nul·la, les variàncies de totes dues poblacions són iguals.

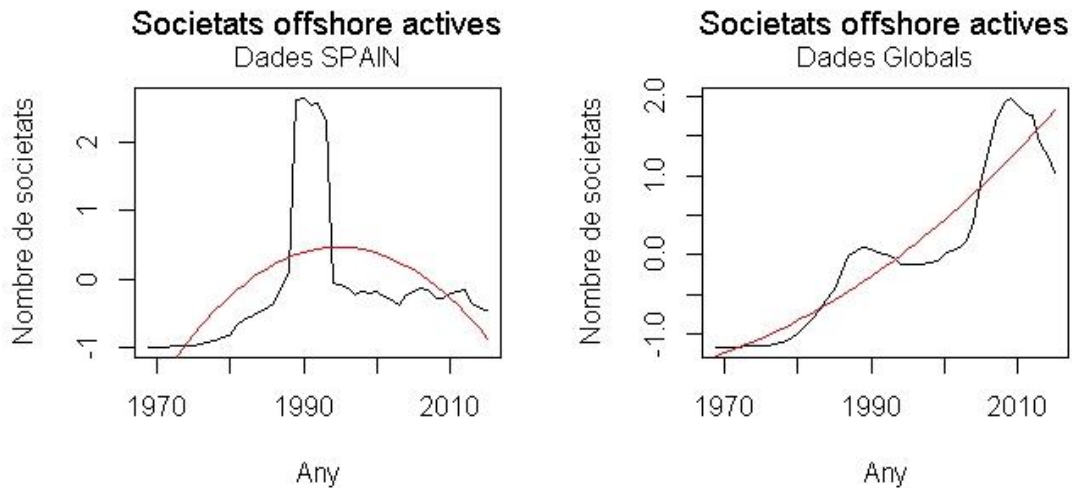
En base al test d'hipòtesis i al plot, concloem que totes dues poblacions es comporten de la mateixa manera al paradís fiscal Niue pel què fa al nombre de societats actives.

```
F test to compare two variances

data: lm1 and lm2
F = 0.8164, num df = 18, denom df = 19, p-value = 0.6706
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3206972 2.1033967
sample estimates:
ratio of variances
 0.8164013
```

4.6.3 Paradís fiscal 3: Panamà

El comportament de totes dues poblacions es pot observar al plot que no és igual, destaca la influència dels outliers de les societats creades i desactivades, per tant gairebé podem concloure que les dues poblacions no es comporten igual respecte l'evolució anual del nombre de societats actives.



Gràfic 21. Nombre de societats actives al paradís fiscal Panamà. Dades per a SPAIN i globals (dreta). Font: Elaboració pròpia.

El model utilitzat per la població de societats espanyoles és un model d'ordre 2

```
Call:
lm(formula = scale(pfsp$acti) ~ poly(as.numeric(pfsp$data), n,
  raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.6902 -0.5590 -0.4182  0.2116  2.2515

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.296e+04  3.916e+03  -3.310  0.00209 **
poly(as.numeric(pfsp$data), n, raw = TRUE)1  1.299e+01  3.927e+00   3.309  0.00209 **
poly(as.numeric(pfsp$data), n, raw = TRUE)2 -3.257e-03  9.844e-04  -3.309  0.00210 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9014 on 37 degrees of freedom
Multiple R-squared:  0.2291,    Adjusted R-squared:  0.1874
F-statistic: 5.497 on 2 and 37 DF,  p-value: 0.008125
```

S'ha obtingut un model amb tots els paràmetres significatius, un coeficient de determinació de 0,18 i un p-valor de 0,008 resultant de l'estadístic F igual a 5,49 amb 2 i 37 g.ll., no és un bon model degut a la influència del pic de societats actives.

Pel que fa al model utilitzat per la població global, és també un model d'ordre 2.

```
Call:
lm(formula = scale(pffs$acti) ~ poly(as.numeric(pffs$data), n2,
  raw = TRUE))

Residuals:
    Min       1Q   Median       3Q      Max
-0.62210 -0.16053 -0.02029  0.21955  0.69394

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.088e+03  1.196e+03   3.419  0.001366 **
poly(as.numeric(pffs$data), n2, raw = TRUE)1 -4.175e+00  1.201e+00  -3.475  0.001160 **
poly(as.numeric(pffs$data), n2, raw = TRUE)2  1.065e-03  3.017e-04   3.531  0.000984 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3401 on 44 degrees of freedom
Multiple R-squared:  0.8893,    Adjusted R-squared:  0.8843
F-statistic: 176.8 on 2 and 44 DF,  p-value: < 2.2e-16
```

S'ha obtingut un model amb tots els paràmetres significatius, un coeficient de determinació de 0,88 i un p-valor de pràcticament zero, resultat de l'estadístic F igual a 176,8 amb 2 i 44 g.ll., per tant, es pot afirmar que és un model que ajusta molt bé les dades.

El resultat del test dona com a resultat un p-valor molt proper a zero, amb un estadístic de contrast F de 7,02 amb 37 i 44 g.ll, respectivament.

Per tant, tenim prou evidències per rebutjar la hipòtesis nul·la, la variància de les dues poblacions és diferent.

```
F test to compare two variances
```

```
data: 1m1 and 1m2
F = 7.0237, num df = 37, denom df = 44, p-value = 3.897e-09
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 3.783661 13.263763
sample estimates:
ratio of variances
      7.023675
```

Podem concloure que totes dues poblacions no es comporten de manera semblant.

5 CONCLUSIONS

De l'anàlisi realitzada amb el software R per comprovar si el comportament de les societats espanyoles incloses als Papers de Panamà del bufet d'advocats de Mossack Fonseca és igual al de totes les societats incloses als mateixos Papers de Panamà i del mateix bufet, es pot extreure una ràpida conclusió i és que no. El comportament de les societats espanyoles als papers de Panamà no és igual al del global de societats.

En gairebé cap dels conceptes claus analitzats, s'ha trobat un indicador que reflecteixi que es comporten de la mateixa manera, de fet tenim evidències per afirmar que es comporten de manera diferent.

Respecte al nombre de societats creades anualment, tot i realitzar dos estudis per separat, degut a què hi ha una dada extrema a la població de les societats espanyoles, totes dues anàlisi han donat un resultat negatiu a la hipòtesi inicial de que totes dues poblacions es comportaven igual. Respecte al nombre de societats desactivades anualment, ens trobem amb la mateixa característica que amb les societats creades, davant la presència d'un outlier s'han realitzat dues anàlisi diferents, una incloent-ho i l'altra sense ell, i de la mateixa manera totes dues donen un resultat negatiu a la hipòtesi de que les dues poblacions es comporten igual.

Respecte al nombre de societats actives anualment, també s'ha extret un resultat negatiu a la hipòtesis que les dues poblacions es comportin igual. Tot i així, al temps mitjà de vida de les societats, tenint en compte només les societats que han estat creades i desactivades en algun moment, el resultat no ha estat tan clar com en els tres apartats anteriors, gràficament totes dues poblacions tenen un comportament molt semblant i és en el contrast d'hipòtesis on, amb un nivell de significació de 0,05, s'ha rebutjat la hipòtesi de que totes dues poblacions tenen un comportament igual. Si el nivell de confiança escollit fos al 99%, no tindríem prou evidències per rebutjar aquesta hipòtesi.

Finalment, respecte a la comparació tant de l'evolució de les societats actives com de les societats creades, només al paradís fiscal de Niue s'ha pogut acceptar la hipòtesi de que totes dues poblacions tenen el mateix comportament. S'ha de tenir en compte que només s'ha pogut realitzar la comparació de tres paradisos fiscals, ja que de la resta no hi ha suficients dades per a que la modelització sigui representativa de les mateixes.

Finalment, com a conclusió global a la pregunta de si el comportament de les societats espanyoles incloses als “Papers de Panamà”, del bufet d’advocats de Mossack Fonseca, és igual al de totes les societats incloses als mateixos, la resposta és no, són diferents.

Com a continuació del treball realitzat seria interessant trobar el motiu de per què totes dues poblacions es comporten de manera diferent: quines causes hi ha per l’outlier en les dades pel paradís fiscal de Panamà, per què les societats espanyoles comencen a decreïxer molt abans que les societats globals, i un llarg etcètera.

BIBLIOGRAFIA

Obermayer, Bastian; Obermaier, Frederik. *LOS PAPELES DE PANAMÁ*. 1a ed. Cortés Fernández, Lara; Guelbunzu de San Eustaquio, Ana; Viejo Pérez, María José; Álvarez Grifoli, Lidia. Península, 2016. 464p. ISBN 978-8499425344.

The International Consortium of Investigative Journalists. *Offshore Leaks Database* [en línea]. Disponible en: <https://offshoreleaks.icij.org/>

R-Project. *The Comprehensive R Archive Network* [en línea]. Disponible en: <https://cran.r-project.org/>

ANNEX

Lectura de les dades

```
#####  
##BLOC 1. ANÀLISI DESCRIPTIU I LECTURA DE DADES ##  
#####  
  
adre=read.csv("Addresses.csv", header = T, sep="," , fill =  
T)  
edges=read.csv("all_edges.csv", header = T, sep="," , fill =  
T)  
enti=read.csv("Entities.csv", header = T, sep="," , fill =  
T)  
inter=read.csv("Intermediaries.csv", header = T, sep="," ,  
fill = T)  
offi=read.csv("Officers.csv", header = T, sep="," , fill =  
T)  
  
#install.packages("sqldf")  
library(sqldf)  
  
spain=sqldf('select * from enti where countries="Spain"')  
table(spain$service_provider)  
spain[c(spain$service_provider=="Portcullis Trustnet"),]  
spain=spain[-1167,]  
  
spain_inter=sqldf('  
    select spain.name AS enti, inter.*  
    from inter, spain, edges  
    where spain.node_id=node_2  
    and inter.node_id=edges.node_1;')  
  
spain_offi=sqldf('  
    select spain.name AS enti, offi.*  
    from offi, spain, edges  
    where spain.node_id=node_2  
    and offi.node_id=edges.node_1;')  
  
spain_adre=sqldf('  
    select spain.name, adre.*  
    from adre, spain, edges  
    where spain.node_id=node_1  
    and adre.node_id=edges.node_2  
    and rel_type like "%registered  
address%";')
```



```

spain_adre2=sqldf('
    select spain_offi.name, adre.*
    from adre, spain_offi, edges
    where spain_offi.node_id=node_1
    and adre.node_id=edges.node_2
    and rel_type like "%registered
address%";')

##Dades Fonseca
fonseca=sqldf('select * from enti where
service_provider="Mossack Fonseca"')

fonseca_inter=sqldf('
    select fonseca.name AS enti,inter.*
    from inter, fonseca, edges
    where fonseca.node_id=node_2
    and inter.node_id=edges.node_1;')

fonseca_offi=sqldf('
    select fonseca.name AS enti, offi.*
    from offi, fonseca, edges
    where fonseca.node_id=node_2
    and offi.node_id=edges.node_1;')

fonseca_adre=sqldf('
    select fonseca.name, adre.*
    from adre, fonseca, edges
    where fonseca.node_id=node_1
    and adre.node_id=edges.node_2
    and rel_type like "%registered
address%";')

fonseca_adre2=sqldf('
    select fonseca_offi.name, adre.*
    from adre, fonseca_offi, edges
    where fonseca_offi.node_id=node_1
    and adre.node_id=edges.node_2
    and rel_type like "%registered
address%";')

```

Anàlisi descriptiu de les dades

```
#Anàlisi descriptiu

sx1=spain$incorporation_date
sx1=as.character(sx1)
sx1=strsplit(sx1, "-")
sx1=matrix(unlist(sx1), ncol=3, byrow=TRUE)

a=NULL
b=NULL
for(i in 1:dim(spain)[1])
{
  a=as.character(spain$incorporation_date[i])
  b=as.character(spain$inactivation_date[i])
  a=unlist(strsplit(a, "-"))
  b=unlist(strsplit(b, "-"))
  spain$acti[i]=a[3]
  spain$deac[i]=b[3]
}

a=NULL
b=NULL
for(i in 1:dim(fonseca)[1])
{
  a=as.character(fonseca$incorporation_date[i])
  b=as.character(fonseca$inactivation_date[i])
  a=unlist(strsplit(a, "-"))
  b=unlist(strsplit(b, "-"))
  fonseca$acti[i]=a[3]
  fonseca$deac[i]=b[3]
  print(i)
}

#install.packages("pastecs")
library(pastecs)
options(scipen=10)

#Any creació societat

aspain=as.data.frame(stat.desc(spain$acti))
afonseca=as.data.frame(stat.desc(fonseca$acti))
View(fonseca)

#Any desactivació societat

dspain=as.data.frame(stat.desc(spain$deac))
dfonseca=as.data.frame(stat.desc(fonseca$deac))
dfonseca=as.data.frame(stat.desc(fonseca$deac[fonseca$deac!
=1930 | fonseca$deac!=0]))
```

```

#Paradisos fiscals
spain$jurisdiction_description=as.character(factor(spain$jurisdiction_description))
fonseca$jurisdiction_description=as.character(factor(fonseca$jurisdiction_description))
jspain=as.data.frame(t(table(spain$jurisdiction_description)))
jspain=jspain[-1]
jspain=jspain[order(-jspain$Freq),]
jfonseca=as.data.frame(t(table(fonseca$jurisdiction_description)))
jfonseca=jfonseca[-1]
jfonseca=jfonseca[order(-jfonseca$Freq),]

##OFFI
View(spain_offi)

ospain=as.data.frame(t(table(spain_offi$name)))
ospain=ospain[-1]
ospain=ospain[order(-ospain$Freq),]
ofonseca=as.data.frame(t(table(fonseca_offi$name)))
ofonseca=ofonseca[-1]
ofonseca=ofonseca[order(-ofonseca$Freq),]

View(spain_offi)
options(scipen=5)

```

Creació dels models i comparació del comportament de les dues poblacions.

Societats creades.

```
#####  
##BLOC 2. CONTRAST HIPOTESIS COMPORTAMENT ##  
#####  
  
#Actives spain  
  
sx1=spain$incorporation_date  
sx1=as.character(sx1)  
sx1=strsplit(sx1, "-")  
sx1=matrix(unlist(sx1), ncol=3, byrow=TRUE)  
  
join_sp =NULL  
for(i in 1:dim(sx1)[1])  
{  
  join_sp[i]=sx1[i,3]  
}  
  
#Creamos los data.frame para cada poblacion  
data_spain=sort(unique(join_sp))  
count_spain=as.vector(table(join_sp))  
spainf=data.frame(data_spain,count_spain)  
  
#Mossack Fonseca Entities  
  
fx1=fonseca$incorporation_date  
fx1=as.character(fx1)  
fx1=strsplit(fx1, "-")  
fx1=matrix(unlist(fx1), ncol=3, byrow=TRUE)  
  
join_fs =NULL  
for(i in 1:dim(fx1)[1])  
{  
  join_fs[i]=fx1[i,3]  
  print(i)  
}  
  
data_fons=sort(unique(join_fs))  
count_fons=as.vector(table(join_fs))  
fonsf = data.frame(data_fons,count_fons)  
fonsf=fonsf[-c(1:17),]  
fonsf$data_fons=as.character(factor(fonsf$data_fons))
```

```

#Models
lm1=lm(scale(count_spain) ~ poly(as.numeric(data_spain),2,
raw=TRUE))
lm2=lm(scale(fonsf$count_fons) ~
poly(as.numeric(fonsf$data_fons),3, raw=TRUE))
summary(lm1)
summary(lm2)

anova(lm1)

par(mfrow=c(1,2))
plot(data_spain, scale(count_spain), type="l",main="Creació
de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades SPAIN"
mtext(subtitle,side=3,line=0.5, font =1)
lines(data_spain, predict(lm1), col="red")
plot(fonsf$data_fons, scale(fonsf$count_fons),
type="l",main="Creació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades Globals"
mtext(subtitle,side=3,line=0.5, font =1)
lines(fonsf$data_fons, predict(lm2), col="red")

var.test(lm1,lm2)

##Anàlisis sense l'outlier a Spain

count_spain[14] #outlier

lm1=lm(scale(count_spain[-14]) ~
poly(as.numeric(data_spain[-14]),3, raw=TRUE))
plot(data_spain[-14], scale(count_spain[-14]),
type="l",main="Creació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades SPAIN"
mtext(subtitle,side=3,line=0.5, font =1)
lines(data_spain[-14], predict(lm1), col="red")
plot(fonsf$data_fons, scale(fonsf$count_fons),
type="l",main="Creació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades Globals"
mtext(subtitle,side=3,line=0.5, font =1)
lines(fonsf$data_fons, predict(lm2), col="red")

summary(lm1)
summary(lm2)
var.test(lm1,lm2)

```

Societats desactivades.

```
#Societats inactives

#Espanya

dx1=spain$inactivation_date
dx1=as.character(dx1)
dx1=strsplit(dx1, "-")
dx1=matrix(unlist(dx1), ncol=3, byrow=TRUE)

deac_sp =NULL
for(i in 1:dim(dx1)[1])
{
  deac_sp[i]=dx1[i,3]
}

data_spain2=sort(unique(deac_sp))
count_spain2=as.vector(table(deac_sp))
spainf2 = data.frame(data_spain2,count_spain2)

##Mossack Fonseca

fx2=fonseca$inactivation_date
fx2=as.character(fx2)
fx2=strsplit(fx2, "-")
fx2=matrix(unlist(fx2), ncol=3, byrow=TRUE)

deac_fs =NULL
for(i in 1:dim(fx2)[1])
{
  deac_fs[i]=fx2[i,3]
  print(i)
}

data_fons=sort(unique(deac_fs))
count_fons=as.vector(table(deac_fs))
fonsf2=data.frame(data_fons,count_fons)
fonsf2=fonsf2[-c(1:9),]
fonsf2$data_fons=as.character(factor(fonsf2$data_fons))

#Models
lm1=lm(scale(count_spain2) ~
poly(as.numeric(data_spain2),2, raw=TRUE))
lm2=lm(scale(fonsf2$count_fons) ~
poly(as.numeric(fonsf2$data_fons),1, raw=TRUE))
summary(lm1)
summary(lm2)
```

```

par(mfrow=c(1,2))
plot(data_spain2, scale(count_spain2),
type="l",main="Desactivació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades SPAIN"
mtext(subtitle,side=3,line=0.5, font =1)
lines(data_spain2, predict(lm1), col="red")
plot(fonsf2$data_fons, scale(fonsf2$count_fons),
type="l",main="Desactivació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades Globals"
mtext(subtitle,side=3,line=0.5, font =1)
lines(fonsf2$data_fons, predict(lm2), col="red")

var.test(lm1,lm2)

#ANALISIS SIN EL OUTLIER [7]
#SPAIN ANALISIS REGRESION
count_spain2[7] #outlier

lm1=lm(scale(count_spain2[-7]) ~
poly(as.numeric(data_spain2[-7]),2, raw=TRUE))
summary(lm1)

#Comparació
par(mfrow=c(1,2))
plot(data_spain2[-7], scale(count_spain2[-7]),
type="l",main="Desactivació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades SPAIN"
mtext(subtitle,side=3,line=0.5, font =1)
lines(data_spain2[-7], predict(lm1), col="red")
plot(fonsf2$data_fons, scale(fonsf2$count_fons),
type="l",main="Desactivació de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades Globals"
mtext(subtitle,side=3,line=0.5, font =1)
lines(fonsf2$data_fons, predict(lm2), col="red")

var.test(lm1,lm2)

```

Societats actives.

```
##ACTIVE THROUGH THE YEARS##

colnames(spainf2)=c("data_spain", "count_spain")
spainf3=merge(spainf,spainf2,by=c("data_spain"), all =T)

for(i in 1:dim(spainf3)[1])
{
  if(is.na(spainf3$count_spain.y[i]))
  {
    spainf3$count_spain.y[i]=0
  }
}

fonsf3=merge(fonsf,fonsf2,by=c("data_fons"), all =T)
for(i in 1:dim(fonsf3)[1])
{
  if(is.na(fonsf3$count_fons.y[i]))
  {
    fonsf3$count_fons.y[i]=0
  }
}

#Creem el nou atribut amb les societats actives cada any

spainf3$acti[1]=spainf3$count_spain.x[1]-
spainf3$count_spain.y[1]
for(i in 2:dim(spainf3)[1])
{
  spainf3$acti[i]=spainf3$acti[i-
1]+spainf3$count_spain.x[i]-spainf3$count_spain.y[i]
}

fonsf3$acti[1]=fonsf3$count_fons.x[1]-
fonsf3$count_fons.y[1]
for(i in 2:dim(fonsf3)[1])
{
  fonsf3$acti[i]=fonsf3$acti[i-1]+fonsf3$count_fons.x[i]-
fonsf3$count_fons.y[i]
}

#Es procedeix a la creació dels models

spainf3$data_spain=as.character(spainf3$data_spain)
fonsf3$data_fons=as.character(factor(fonsf3$data_fons))

lm1=lm(scale(spainf3$acti) ~
poly(as.numeric(spainf3$data_spain),3,raw=TRUE))
```



```

lm2=lm(scale(fonsf3$acti) ~
poly(as.numeric(fonsf3$data_fons),3,raw=TRUE))
summary(lm1)
summary(lm2)

par(mfrow=c(1,2))
plot(spainf3$data_spain, scale(spainf3$acti),
type="l",main="Societats offshore actives",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades SPAIN"
mtext(subtitle,side=3,line=0.5, font =1)
lines(spainf3$data_spain, predict(lm1), col="red")
plot(fonsf3$data_fons, scale(fonsf3$acti),
type="l",main="Societats offshore actives",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades Globals"
mtext(subtitle,side=3,line=0.5, font =1)
lines(fonsf3$data_fons, predict(lm2), col="red")

var.test(lm1,lm2)

```

Temps mitjà de vida de les societats offshore

```
## Meantime life ##

#Temps mitja de vida de les societats espanyoles sense
tenir en compte les societats que no tenen data de
desactivació#
spain$acti=as.numeric(spain$acti)
spain$deac=as.numeric(spain$deac)
for(i in 1:dim(spain)[1])
{
  if(is.na(spain$deac[i]))
  {
    spain$deac[i]=0
  }
}

for(i in 1:dim(spain)[1])
{
  if(spain$deac[i] > spain$acti[i])
  {
    spain$ltime[i]=spain$deac[i]-spain$acti[i]
  }
  else
  {spain$ltime[i]=NA}
}

spainf4=aggregate(spain$ltime, by=list(acti=spain$acti),
FUN=mean, na.rm=T)
spainf4$x[c(38,40)]=0

#Temps mitja de vida de totes les societats sense tenir en
compte les societats que no tenen data de desactivació#

fonseca$acti=as.numeric(fonseca$acti)
fonseca$deac=as.numeric(fonseca$deac)

for(i in 1:dim(fonseca)[1])
{
  if(is.na(fonseca$deac[i]))
  {
    fonseca$deac[i]=NA
    print(i)
  }
}

for(i in 1:dim(fonseca)[1])
{
  if(is.na(fonseca$deac[i]))
  {
```

```

    {fonseca$lttime[i]=NA}
  }
else if(is.na(fonseca$acti[i]))
{fonseca$lttime[i]=NA}
else if(fonseca$deac[i] > fonseca$acti[i])
{
  fonseca$lttime[i]=fonseca$deac[i]-fonseca$acti[i]
}
else
{fonseca$lttime[i]=NA}
print(i)
}

fonsf4=aggregate(fonseca$lttime, by=list(acti=fonseca$acti),
FUN=mean, na.rm=T)
fonsf4=fonsf4[-c(1:17),]
fonsf4$x[47]=0

##Models##

fonsf4$acti=as.character(factor(fonsf4$acti))
spainf4$acti=as.character(factor(spainf4$acti))

lm1=lm((spainf4$x) ~
poly(as.numeric(spainf4$acti),3,raw=TRUE))
lm2=lm((fonsf4$x) ~
poly(as.numeric(fonsf4$acti),3,raw=TRUE))
summary(lm1)
summary(lm2)

par(mfrow=c(1,2))
plot(spainf4$acti, (spainf4$x), type="l",main="Temps mitjà
de vida",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades SPAIN"
mtext(subtitle,side=3,line=0.5, font =1)
lines(spainf4$acti, predict(lm1), col="red") #ajustamos una
regresion cúbica porque es la que mejor ajusta
plot(fonsf4$acti, (fonsf4$x), type="l",main="Temps mitjà de
vida",
      xlab="Any", ylab="Nombre de societats",font.main = 1)
subtitle="Dades Globals"
mtext(subtitle,side=3,line=0.5, font =1)
lines(fonsf4$acti, predict(lm2), col="red") #ajustamos una
regresion cúbica porque es la que mejor ajusta

var.test(lm1,lm2)

```

Comparació entre els paradisos fiscals

```
## Test per als diferents Paradisos fiscals##

spain$jurisdiction_description=factor(spain$jurisdiction_de
scription)
table(spain$jurisdiction_description)

###Funció per calcular Societats actives per any d'un
paradis fiscal###

dframe_cre=function(incorp)
{
  data=sort(unique(incorp$acti))
  count=as.vector(table(incorp$acti))
  dframe = data.frame(data,count)

  data=sort(unique(incorp$deac))
  count2=as.vector(table(incorp$deac))
  dframe2=data.frame(data,count2)

  dframe3=merge(dframe,dframe2,by=c("data"), all =T)

  for(i in 1:dim(dframe3)[1])
  {
    if(is.na(dframe3$count[i]))
    {
      dframe3$count[i]=0
    }
    else if(is.na(dframe3$count2[i]))
    {
      dframe3$count2[i]=0
    }
  }

  if(dframe3$data[1]==0)
  {dframe3=dframe3[-c(dframe3$data==0),]}

  dframe3$data=as.character(factor(dframe3$data))

  dframe3$acti[1]=dframe3$count[1]-dframe3$count2[1]
  for(i in 2:dim(dframe3)[1])
  {
    dframe3$acti[i]=dframe3$acti[i-1]+dframe3$count[i]-
dframe3$count2[i]
  }
  return(dframe3)
}
```

```

##Paradisos fiscals on hi ha societats offshore
espanyoles##

sb=dframe_cre(spain[spain$jurisdiction_description=="Bahamas",])
sbv=dframe_cre(spain[spain$jurisdiction_description=="British Virgin Islands",])
sni=dframe_cre(spain[spain$jurisdiction_description=="Niue",])
sp=dframe_cre(spain[spain$jurisdiction_description=="Panama",])
ss=dframe_cre(spain[spain$jurisdiction_description=="Samoa",])
sse=dframe_cre(spain[spain$jurisdiction_description=="Seychelles",])

##Paradisos fiscals on hi ha societats offshore (mateixos paradisos fiscals que tenen societats espanyoles)##
fb=dframe_cre(fonseca[fonseca$jurisdiction_description=="Bahamas",])
fbv=dframe_cre(fonseca[fonseca$jurisdiction_description=="British Virgin Islands",])
fni=dframe_cre(fonseca[fonseca$jurisdiction_description=="Niue",])
fp=dframe_cre(fonseca[fonseca$jurisdiction_description=="Panama",])
fs=dframe_cre(fonseca[fonseca$jurisdiction_description=="Samoa",])
fse=dframe_cre(fonseca[fonseca$jurisdiction_description=="Seychelles",])

```

Comparació del nombre de societats creades per paradís fiscal

```
##Comparacio nombre societats creades per any##
comparacio=function(pfsp,pffs,n,n2)
{
  par(mfrow=c(1,2))
  lm1=lm(scale(pfsp$count) ~
poly(as.numeric(pfsp$data),n,raw=TRUE))
  plot(as.character(pfsp$data), scale(pfsp$count),
type="l",main="Creació3 de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main =
1)
  subtitle="Dades SPAIN"
  mtext(subtitle,side=3,line=0.5, font =1)
  lines(pfsp$data, predict(lm1), col="red")
  lm2=lm(scale(pffs$count) ~
poly(as.numeric(pffs$data),n2,raw=TRUE))
  plot(as.character(pffs$data), scale(pffs$count),
type="l",main="Creació3 de societats offshore",
      xlab="Any", ylab="Nombre de societats",font.main =
1)
  subtitle="Dades Globals"
  mtext(subtitle,side=3,line=0.5, font =1)
  lines(pffs$data, predict(lm2), col="red")
  summary(lm2)
  summary(lm1)
  var.test(lm1,lm2)
  return(list(summary(lm1),summary(lm2),var.test(lm1,lm2)))
}

#Bahames
fb=fb[-c(1:5),]
fb$data=as.character(factor(fb$data))
comparacio(sb,fb,2,2)

#Model no representatiu per no tenim suficients dades

#British Virgin Islands
fbv=fbv[-c(1:10),]
fbv$data=as.character(factor(fbv$data))
comparacio(sbv,fbv,2,2)
#ajustem una regresion cuadrática porque es la que millor
ajusta

#Niue
for(i in 1:6)
{
  sni[nrow(sni) + 1,] = c(2009+i,0,0,9)
}
fni$data=as.character(factor(fni$data))
```

```

fni[nrow(fni) + 1,] = c(2015,0,0,1009)

comparacio(sni,fni,1,1)

#ajustem una regresion lineal porque es la que millor
ajusta

#Panama sense l'outlier
sp=sp[-14,]
fp=fp[-c(1:17,65),]
fp$data=as.character(factor(fp$data))
comparacio(sp,fp,2,1)
#Model no significatiu

#Samoa
fs=fs[-c(1:7),]
fs$data=as.character(factor(fs$data))
comparacio(ss,fs,1,1)
#Model no significatiu, no representatiu per no tenir
suficients dades

#Seychelles
fse=fse[-c(1:9),]
fse$data=as.character(factor(fse$data))
comparacio(sse,fse,1,2)
#Model no significatiu, no representatiu per no tenir
suficients dades

```

Comparació del nombre de societats actives per paradís fiscal

```
##Comparació3 Nombre de societats actives per any##

comparacio2=function(pfsp,pffs,n,n2)
{
  par(mfrow=c(1,2))
  lm1=lm(scale(pfsp$acti) ~
poly(as.numeric(pfsp$data),n,raw=TRUE))
  plot(as.character(pfsp$data), scale(pfsp$acti),
type="l",main="Societats offshore actives",
      xlab="Any", ylab="Nombre de societats",font.main =
1)
  subtitle="Dades SPAIN"
  mtext(subtitle,side=3,line=0.5, font =1)
  lines(pfsp$data, predict(lm1), col="red")
  lm2=lm(scale(pffs$acti) ~
poly(as.numeric(pffs$data),n2,raw=TRUE))
  plot(as.character(pffs$data), scale(pffs$acti),
type="l",main="Societats offshore actives",
      xlab="Any", ylab="Nombre de societats",font.main =
1)
  subtitle="Dades Globals"
  mtext(subtitle,side=3,line=0.5, font =1)
  lines(pffs$data, predict(lm2), col="red")
  summary(lm2)
  summary(lm1)
  var.test(lm1,lm2)
  return(list(summary(lm1),summary(lm2),var.test(lm1,lm2)))
}

#Bahames

comparacio2(sb,fb,1,2)
#Model no significatiu

#British Virgin Islands

comparacio2(sbv,fbv,2,2)
#ajustem una regresion cuadrática porque es la que millor
ajusta

#Niue

comparacio2(sni,fni,2,2)

#ajustem una regresion cuadrática porque es la que millor
ajusta
```



```
#Panama
fp=dframe_cre(fonseca[fonseca$jurisdiction_description=="Pa
nama",])
fp=fp[-c(1:17,65),]
fp$data=as.character(factor(fp$data))
comparacio2(sp,fp,2,2)
```

```
#Samoa
```

```
comparacio2(ss,fs,2,2)
#Model no significatiu
```

```
#Seychelles
```

```
comparacio2(sse,fse,1,1)
#Model no explicatiu, no hi ha suficients dades
```