

Grau en Estadística

Títol: Estudi de la resistència als antibiòtics a la Unió Europea utilitzant MFA.

Autor: Guillem Querol Lloveria

Director: Antonio Miñarro Alonso
Marta Cubedo Cullere

Departament: Departament de Genètica, Microbiologia i Estadística

Convocatòria: Gener 2021



Agraïments

En primer lloc, m'agradaria agrair tant a l'Antonio com a la Marta, l'esforç i el temps que han dedicat en ajudar-me, ja que aquest treball ha estat possible gràcies a la seva ajuda.

També voldria agrair, tant els meus pares com el meu germà, el suport que he rebut i la paciència que han mostrat en tot moment.

Resum

Els resultats presentats en l'informe extret de la European Centre for Disease Prevention and Control (ECDC) estan basats en dades de resistència antimicrobiana (RAM) d'aïllats invasius comunicats a la xarxa europea de vigilància de la resistència antimicrobiana (EARS-Net) per 30 països de la Unió Europea (UE) i de l'Espai Econòmic Europeu (EEE) el 2019 (dades referent al 2018), i en anàlisis de tendències de les dades reportades pels països participants per al període 2015-2018.

El projecte *Surveillance of antimicrobial resistance in Europe 2018* és un estudi que té com a objectiu mostrar la RAM que hi ha a Europa, mostrant les variacions que hi ha entre els diferents països, diferents bacteris i antibiòtics.

En aquest estudi, a través de diferents tècniques d'Anàlisis Multivariant, s'analitza la resistència dels diferents tipus de bacteris a la combinació d'antibiòtics que s'ha observat en els diferents països seleccionats i l'evolució que han tingut aquests tractaments.

Paraules Clau

Resistència Antimicrobiana (RAM), *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Acinetobacter species*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Enterococci*, Anàlisis de Component Principals (PCA), Anàlisis Factorial Múltiple (MFA).

Clasificación AMS

- 62P10 Applications to biology and medical sciences
- 62H25 Factor analysis and principal components; correspondence analysis
- 62-07 Data analysis

Abstract

The results presented in the report extracted from the European Center for Disease Prevention and Control (ECDC) are based on antimicrobial resistance (RAM) data of invasive isolates reported to the European Antimicrobial Resistance Surveillance Network (EARS Net) by 30 countries of the European Union (EU) and the European Economic Area (EEA) in 2019 (data referring to 2018), and in analysis of trends in the data reported by the participating countries for the period 2015-2018.

The *Surveillance of antimicrobial resistance in Europe 2018* project is a study that aims to show the RAM that exists in Europe, showing the variations that exist between different countries, different bacteria and antibiotics.

In this study, through different Multivariate Analysis techniques, the resistance of different types of bacteria to the combination of antibiotics that can be observed in the different selected countries and the evolution that these have had will be analyzed.

Keywords

Antimicrobial Resistance (RAM), *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Acinetobacter species*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Enterococci*, Principal Component Analysis (PCA), Multiple Factor Analysis (MFA).

Índex

1.	Introducció	1
1.1.	Justificació.....	1
1.2.	Objectius.....	2
1.3.	Metodologia.....	2
1.4.	Estructura.....	3
1.5.	Descripció de la base de dades	3
2.	Introducció a la Microbiologia	7
2.1.	Què és un bacteri?	7
2.2.	Tipus de bacteris.....	7
2.3.	Bacteris que s'estudiaran en aquest treball.....	8
2.3.1.	<i>Escheríchia coli</i>	8
2.3.2.	<i>Klebsiella pneumoniae</i>	10
2.3.3.	<i>Pseudomonas aeruginosa</i>	11
2.3.4.	<i>Acinetobacter spp.</i>	13
2.3.5.	<i>Streptococcus pneumoniae</i>	14
2.3.6.	<i>Staphylococcus aureus</i>	16
2.3.7.	<i>Enterococci</i>	17
3.	Anàlisi de Components Principals	20
3.1.	Què és l'Anàlisi de Components Principals?	20
3.2.	Per què l'anàlisi de components principals?.....	21
3.3.	Anàlisi de Components Principals en R	22
3.3.1.	Components principals 1 vs 2.....	26
3.3.2.	Components principals 1 vs 3.....	29
3.3.3.	Components principals 2 vs 3.....	32
3.4.	Interpretació del PCA	34
4.	Anàlisi Factorial Múltiple	37
4.1.	Anàlisi Factorial Múltiple en R.....	38
4.1.1.	Explicació de la variabilitat en cada dimensió.....	39
4.1.2.	Contribució de cada grup a les dimensions.....	40
4.1.3.	Contribució de cada individu a les dimensions	44
4.1.4.	Contribució de cada variable a les dimensions	48
4.2.	Qualitat de representació en les principals dimensions	52
4.3.	Resultats del MFA.....	59
5.	Conclusió.....	61

6. Bibliografia i Webgrafia.....	62
7. Glossari.....	65

Índex de figures

Figura 1: Boxplot de la resistència d' <i>Escheríchia coli</i> als antibiòtics estratificat per anys	9
Figura 2: Boxplot de la resistència de <i>Klebsiel-la pneumoniae</i> als antibiòtics estratificat per anys.....	11
Figura 3: Boxplot de la resistència de <i>Pseudomonas aeruginosa</i> als antibiòtics estratificat per anys	12
Figura 4: Boxplot de la resistència de <i>Acinetobacter species</i> als antibiòtics estratificat per anys	14
Figura 5: Boxplot de la resistència de <i>Streptococcus pneumoniae</i> als antibiòtics estratificat per anys...	15
Figura 6: Boxplot de la resistència de <i>Staphylococcus aureus</i> als antibiòtics estratificat per anys.....	16
Figura 7: Boxplot de la resistència de <i>Enterococcus faecalis</i> als antibiòtics estratificat per anys	18
Figura 8: Boxplot de la resistència de <i>Enterococcus faecium</i> als antibiòtics estratificat per anys	18
Figura 9: Proporció de la variància explicada acumulada	25
Figura 10: Puntuacions del PC1 enfront PC2	27
Figura 11: Projeccions de les variables per a PC1 i PC2	28
Figura 12: Puntuacions del PC1 enfront PC3	29
Figura 13: Projeccions de les variables per a PC1 i PC3	31
Figura 14: Puntuacions del PC2 enfront PC3	32
Figura 15: Projeccions de les variables per a PC2 i PC3	33
Figura 16: Percentatge de la variabilitat explicat per a cada dimensió del MFA.....	39
Figura 17: Contribució de cada any en la dimensió 1	41
Figura 18: Contribució de cada any en la dimensió 2	41
Figura 19: Contribució de cada any en les dimensions 1 i 2	42
Figura 20: Representació dels grups en dues dimensions	43
Figura 21: Distribució dels països per a les dues dimensions	45
Figura 22: Evolució dels països en els anys per a les dues dimensions	46
Figura 23: Evolucions particulars en els anys per a les dues dimensions	46
Figura 24: Contribució de les variables quantitatives en la dimensió 1	48
Figura 25: Contribució de les variables quantitatives en la dimensió 2	48
Figura 26: Contribució de les variables quantitatives en les dimensions 1 i 2.....	49
Figura 27: Gràfic dels països amb els \cos^2	52
Figura 28: Gràfic de les variables amb $\cos^2 >0.5$	53
Figura 29: Gràfic de les variables amb $\cos^2 >0.75$	53
Figura 30: \cos^2 per variables en la dimensió 1	55
Figura 31: \cos^2 per cada variable en la dimensió 1	56
Figura 32: \cos^2 per variables en la dimensions 1 i 2	56

Índex de taules

Taula 1: Descriptiva any 2015.....	4
Taula 2: Descriptiva any 2016.....	5
Taula 3: Descriptiva any 2017.....	5
Taula 4: Descriptiva any 2018.....	6
Taula 5: Matriu de Correlacions any 2018.....	21
Taula 6: vectors propis per a els 3 components principals.....	23
Taula 7: Proporció de la variància per cada component.....	24
Taula 7: Matriu de components per a les 3 components principals.....	34
Taula 8 : Proporció de la variància de les dimensions amb més importància.....	39
Taula 9: Matriu amb els resultats dels diferents anys.....	40
Taula 10 : Matriu amb els resultats dels diferents països.....	45
Taula 11: Contribucions de les variables quantitatives en les dimensions 1 i 2.....	49
Taula 12: Correlacions entre variables i dimensions.....	50
Taula 13: Cos^2 per variable.....	57
Taula 14: cos^2 per individu.....	58

1. Introducció

La resistència antimicrobiana (RAM) és la capacitat que té un microorganisme per resistir l'acció d'un o més diferents agents antimicrobians. La RAM és considerada una de les principals amenaces per a la salut pública a nivell europeu¹ i mundial², amb un nombre d'infeccions pròxim a les 670.000 persones/any i un nombre de morts al voltant de les 33.000 persones/any³, només en l'Espai Econòmic Europeu (EEE). El cost relacionat per als sistemes sanitaris dels països de la UE / EEE és del voltant de 1.100 milions d'euros⁴.

Aquesta RAM requereix un esforç dintre de cada país per ser controlada i, a la vegada, també una cooperació entre els diferents països.

La resistència adquirida és causada per mutacions de gens bacterians, o adquisició de gens de resistència exògena portats per elements genètics mòbils que es poden estendre horitzontalment entre bacteris. Els bacteris poden adquirir múltiples mecanismes de resistència i, per tant, es tornen resistents a diversos agents antimicrobians, cosa particularment problemàtica, ja que pot limitar greument el tractament disponible alternatiu per a la infecció.

Per a la vigilància d'aquest problema està la Xarxa Europea de Vigilància de la Resistència Antimicrobiana (EARS-Net), encarregada de la vigilància de bacteris que causen infeccions greus. L'objectiu principal de la EARS-Net és analitzar i informar a tots els estats que formen part de la UE/EEE per permetre accions per abordar la RAM.

En aquest treball, s'estudiarà l'evolució de la RAM de diferents bacteris entre els anys 2015 i 2018 i la situació de l'any 2018 dels països que han participat en la EARS-Net amb la finalitat de veure si la situació va millorant amb el pas del temps i quina és la situació en l'últim any del qual tenim dades.

1.1. Justificació

Cal treballar en projectes com aquest, ja que la RAM és una de les principals amenaces de la salut pública, tot i que la gent no és conscient i no es dóna la importància i repercussió que s'hauria de donar. Segons el projecte *Surveillance of antimicrobial resistance in Europe 2018*,

¹ Comissió europea: A European One Health Action Plan against Antimicrobial Resistance

² Organització mundial de la Salut: A European One Health Action Plan against AMR.

³ The Lancet: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(18\)30605-4/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(18)30605-4/fulltext)

⁴ OCDE: Antimicrobial Resistance. Tackling the burden in the European Union.

el cost dels sistemes sanitaris dels països de La UE / EEE és d'uns 1.100 milions d'euros i moren una quantitat estimada de 33.000 persones a l'any.

1.2. Objectius

Els objectius d'aquest treball són:

- Analitzar quina és la situació dels països de la Unió Europea i de l'Espai Econòmic Europeu (EU/EEE) en quant a la resistència d'alguns bacteris patògens als antibiòtics.
- Estudiar quina ha estat l'evolució en els darrers anys, a la Unió Europea, quant a aquesta resistència als antibiòtics.

Per a assolir el primer objectiu tindrem en compte les dades de l'any 2018, últim any en el qual s'ha fet l'estudi i, en els dos objectius, es tindrà en compte la RAM combinada dels diferents antibacterians per a cada bacteri.

1.3. Metodologia

Per aquest treball, primerament, s'han extret les dades del projecte *Surveillance of antimicrobial resistance in Europe 2018*⁵ i, posteriorment, s'ha seleccionat amb quines dades es volia treballar. S'ha decidit treballar amb les taules que presenten la resistència del bacteri enfront del combinat d'antibiòtics, ja que si s'hagués treballat amb totes les taules amb els diferents antibiòtics per a cada bacteri, l'estudi hagués resultat molt laboriós i gran part d'aquestes dades no ens resultarien útils per a complir els objectius esmentats anteriorment.

Un cop introduïdes les dades, s'ha realitzat una estadística descriptiva a través de diferents *boxplots* agrupats per bacteri. Després, s'ha buscat informació dels bacteris i definint les seves característiques i els diferents grups de bacteris. donant especial èmfasi als que es consideren en aquest treball, per a tenir més coneixement sobre aquests.

Un cop les dades s'han depurat i ja s'ha donat una breu informació sobre els coneixements bàsics que necessitem sobre microbiologia, s'ha començat a treballar en l'anàlisi de components principals, utilitzat per a reduir el nombre de dimensions i, així, s'ha pogut treballar amb unes dades menys complexes de les que es tenien en un principi.

⁵ Centre Europeu per a la Prevenció i Control de Malalties (ECDC)

Tot seguit, s'ha dut a terme una sèrie de tècniques d'anàlisi Multivariant. La principal tècnica ha estat l'anàlisi factorial múltiple (MFA). Aquesta tècnica s'utilitza per a estudiar diferents grups de variables al mateix temps.

1.4. Estructura

Al llarg del treball el que s'ha pretès ha estat extreure una informació apta, tant per a gent amb pocs coneixements de bioestadística, com per a un públic amb més coneixements d'aquesta àrea. Per això, el cos del treball s'estructura en 4 seccions principals.

La primera secció és una breu descripció a la microbiologia, més concretament, al camp dels bacteris. La finalitat d'aquesta part ha estat facilitar la informació necessària perquè els lectors puguin entendre que és un bacteri, quins tipus de bacteris hi ha i quines diferències hi ha entre les diferents famílies de bacteris.

La segona part del cos del treball és la part en la qual s'ha dut a terme el PCA, on s'explica en què consisteix i els resultats obtinguts, interpretant els gràfics que s'han obtingut per tal d'assolir un dels objectius del treball, que ha estat saber la situació actual (de l'any 2018) dels diferents països de la UE/EEE.

La tercera secció ha estat el MFA, on s'ha comparat la situació i evolució dels diferents països per tal de veure si hi ha països amb certa relació. S'ha fet 3 comparacions principals. En la primera s'ha comparat els diferents blocs fets en el MFA, on cada bloc era un any. En la segona s'han comparat les diferents variables i després, en la tercera, els diferents països.

La quarta i última part consta de les conclusions que s'han tret després de dur a terme tot el treball.

1.5. Descripció de la base de dades

La base de dades, com ja s'ha dit anteriorment, s'ha tret de l'estudi *Surveillance of antimicrobial resistance in Europe 2018*, estudi públic que es pot trobar a <https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2018.pdf>.

L'estudi original consta de 27 bases de dades diferents, on cada taula té les dades de la resistència d'un bacteri per a un antibiòtic determinat, a diferència de les nostres, on s'ha

treballat amb la resistència d'un bacteri amb el combinat d'antibiòtics i s'ha ajuntat tot en una única taula.

Cada taula de l'estudi original consta de les següents variables:

- COUNTRY: País on s'ha dut terme l'estudi.
- N: Es pot veure el nombre de casos que s'han estudiat en cada país.
- % R: És el percentatge de resistència que hi ha per part del bacteri a l'antibiòtic.
- 95% CI: Percentatge on es concentra la resistència del 95% dels casos de cada país.

Cal destacar que les variables N, % R i 95% CI estan 4 cops, ja que s'han estudiat en els anys 2015, 2016, 2017, 2018.

En el cas de la base de dades amb la que s'ha treballat al llarg d'aquest projecte totes les variables han estat numèriques, ja que la variable COUNTRY s'ha suprimit i s'ha utilitzat el nom dels diferents països com a nom de cada observació. Les variables numèriques que s'han utilitzat ha estat el %R de cada bacteri pels diferents anys, ja que són les dades que es creuen necessàries per a dur a terme aquest treball.

A continuació s'han introduït una sèrie de taules amb la finalitat de veure la distribució de valors en cada any.

2015	Mínim	Q1	Mediana	Mitjana	Q3	Màxim	NA
<i>Echer coli</i>	0.000	3.250	5.500	6.432	7.850	17.100	0
<i>Kleb pneumo</i>	0.000	3.750	17.700	20.950	31.300	59.600	0
<i>Pseudo aeru</i>	0.000	3.700	12.000	14.310	20.450	63.000	0
<i>Acinetobacter</i>	0.000	4.025	25.450	32.504	57.475	87.000	3
<i>Strep pneumo</i>	0.000	2.275	5.550	7.954	11.250	25.000	3
<i>Staphy aureus</i>	0.000	6.550	13.100	17.850	25.000	57.200	0
<i>Enter faecalis</i>	8.600	14.300	31.300	29.180	39.100	49.100	3
<i>Enter faecium</i>	0.000	0.950	10.500	11.310	17.520	45.800	1

Taula 1: Descriptiva any 2015

2016	<i>Mínim</i>	<i>Q1</i>	<i>Mediana</i>	<i>Mitjana</i>	<i>Q3</i>	<i>Màxim</i>	<i>NA</i>
<i>Echer coli</i>	1.100	3.450	5.500	6.623	8.950	22.100	0
<i>Kleb pneumo</i>	0.000	4.500	18.700	21.660	32.300	55.700	0
<i>Pseudo aeru</i>	0.000	4.150	10.700	13.640	19.350	48.900	0
<i>Acinetobacter</i>	0.000	0.450	28.300	33.340	63.250	84.000	4
<i>Strep pneumo</i>	0.000	2.000	4.2000	7.961	10.400	40.000	3
<i>Staphy aureus</i>	1.200	5.450	13.800	16.860	25.550	50.500	0
<i>Enter faecalis</i>	12.500	19.800	33.000	31.380	42.200	56.300	2
<i>Enter faecium</i>	0.000	1.300	8.300	13.550	22.250	46.300	0

Taula 2: *Descriptiva any 2016*

2017	<i>Mínim</i>	<i>Q1</i>	<i>Mediana</i>	<i>Mitjana</i>	<i>Q3</i>	<i>Màxim</i>	<i>NA</i>
<i>Echer coli</i>	1.500	3.400	5.700	6.855	8.800	24.900	0
<i>Kleb pneumo</i>	0.000	5.450	19.400	22.220	32.350	57.100	0
<i>Pseudo aeru</i>	0.000	6.000	10.600	14.480	17.800	59.100	0
<i>Acinetobacter</i>	0.000	1.700	25.200	34.640	73.800	84.300	4
<i>Strep pneumo</i>	0.100	3.000	6.500	10.070	14.800	45.500	2
<i>Staphy aureus</i>	1.000	5.800	12.900	16.090	26.800	44.400	0
<i>Enter faecalis</i>	7.100	18.200	29.700	27.980	36.700	45.900	2
<i>Enter faecium</i>	0.000	1.600	13.300	14.930	27.050	43.900	0

Taula 3: *Descriptiva any 2017*

2018	<i>Mínim</i>	<i>Q1</i>	<i>Mediana</i>	<i>Mitjana</i>	<i>Q3</i>	<i>Màxim</i>	<i>NA</i>
<i>Echer coli</i>	2.000	3.250	4.700	6.468	8.200	19.600	0
<i>Kleb pneumo</i>	0.000	5.200	19.600	21.880	35.450	51.500	0
<i>Pseudo aeru</i>	0.000	3.350	11.000	13.320	19.600	49.400	0
<i>Acinetobacter</i>	0.000	4.150	25.400	33.900	63.770	90.800	3
<i>Strep pneumo</i>	0.100	2.600	5.800	7.200	10.000	26.700	2
<i>Staphy aureus</i>	0.000	6.050	12.100	15.850	25.300	43.000	0
<i>Enter faecalis</i>	6.700	13.400	25.300	25.080	33.700	41.600	2
<i>Enter faecium</i>	0.000	1.950	17.300	17.620	29.700	59.100	0

Taula 4: *Descriptiva any 2018*

També s'ha dut a terme una sèrie de *boxplots*, un per cada bacteri, que es veuen a continuació, en l'apartat on s'introdueix la informació que es considera necessària saber sobre els diferents bacteris.

2. Introducció a la Microbiologia

2.1. Què és un bacteri?

Els bacteris són organismes procariotes unicel·lulars i sense nucli, que es troben en quasi totes els medis i ambients de la Terra. És el tipus de microorganisme més abundant. Moltes espècies també viuen en el cos humà i en el dels animals. Tenim bacteris que habiten a la pell, a la boca, a les vies respiratòries, en el sistema digestiu, al reproductor o al urinari. Hi ha una gran quantitat d'aquests microorganismes dels quals no causen cap dany, fins i tot, són beneficiosos i només uns pocs tipus de bacteris són els que provoquen malalties mitjançant la producció de toxines (substàncies nocives), transmetent, envaint teixits de l'organisme, adherint-se a les cèl·lules de l'hoste o evadint el sistema immunitari. Són els anomenats bacteris patògens⁶.

2.2. Tipus de bacteris

Aquests microorganismes es poden diferenciar per:

- Característiques morfològiques:
 - Cocs: Tenen una forma esfèrica i es poden agrupar en diferents formes.
 - Bacils: Tenen una forma cilíndrica i poden tenir diferents formes als extrems.
 - Espirils: Tenen una forma helicoidal o d'espiral.
 - Altres formes: També existeixen altres formes com l'estrella, quadrada, etc.
- Estructures:
 - Externes: Flagels, pilis, prosteqes, etc.
 - Internes: Endòspores, magnetosomes, carboxisomes, etc.
- Habitat:
 - Temperatura: Els bacteris poden viure en temperatures que van des de 4°C fins a 118°C. Els classifiquem en psicròfils, mesòfils, termòfils.
 - pH: Els bacteris poden créixer en un pH que va des de 0 fins a 11, els classifiquem en acidòfils, neutròfils o basòfils.

⁶ <https://www.msdsalud.es/cuidar-en/infecciones/infecciones-bacterianas/es-una-infeccion-bacteriana.html>

- Medi: Els bacteris poden viure tant en aigua dolça com en aigua salada.
- Moviment:
 - Tactisme: Moviment a través de l'atracció o repulsió a determinades substàncies.
 - Estructures de moviment: Moviment a través dels flagels.
- Resposta a la tinció de Gram: els bacteris responen de dues formes diferents a l'aplicar-les un procés químic anomenat tinció de Gram. Això es deu a la diferent composició de la paret cel·lular del bacteri. Això fa que els tractaments posteriors també siguin diferents.
 - Bacteris grampositius: es tenyeixen de blau a l'aplicar la tinció.
 - Bacteris gramnegatius: es tenyeixen de vermell a l'aplicar la tinció de Gram. Tenen una membrana sota la capsula protectora que els fa resistents a alguns antibiòtics.

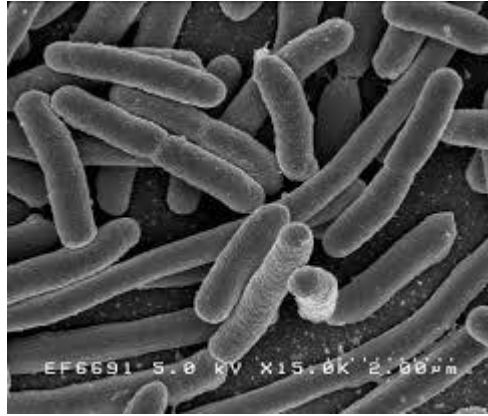
2.3. Bacteris que s'estudiaran en aquest treball

2.3.1. *Escheríchia coli*

El bacteri *Escheríchia coli* és un grup de bacteris gramnegatius que part de la microbiota intestinal que tenen els humans, però algunes de les seves mutacions poden provocar infeccions greus del tub digestiu i de les vies urinàries. Es caracteritza per poder viure en medis i temperatures molt variats. Les infeccions per E. coli es desenvolupen al consumir aliments contaminats, pel contacte amb animals infectats o per la ingestió d'aigua contaminada en una piscina⁷.

Causen diarrea, de vegades greu o amb sang, i dolor abdominal.

⁷ Msdmanuals: infecciones por Escherichia coli



Font: Google

Segons l'informe de la ECDC, és la causa més freqüent d'infeccions del torrent sanguini i infeccions del tracte urinari a la UE / EEE i participa en infeccions d'origen comunitari i sanitari. En aquest cas, la resistència es desenvolupa a través de mutacions o a través de l'adquisició d'elements mòbils que codifiquen mecanismes de resistència.

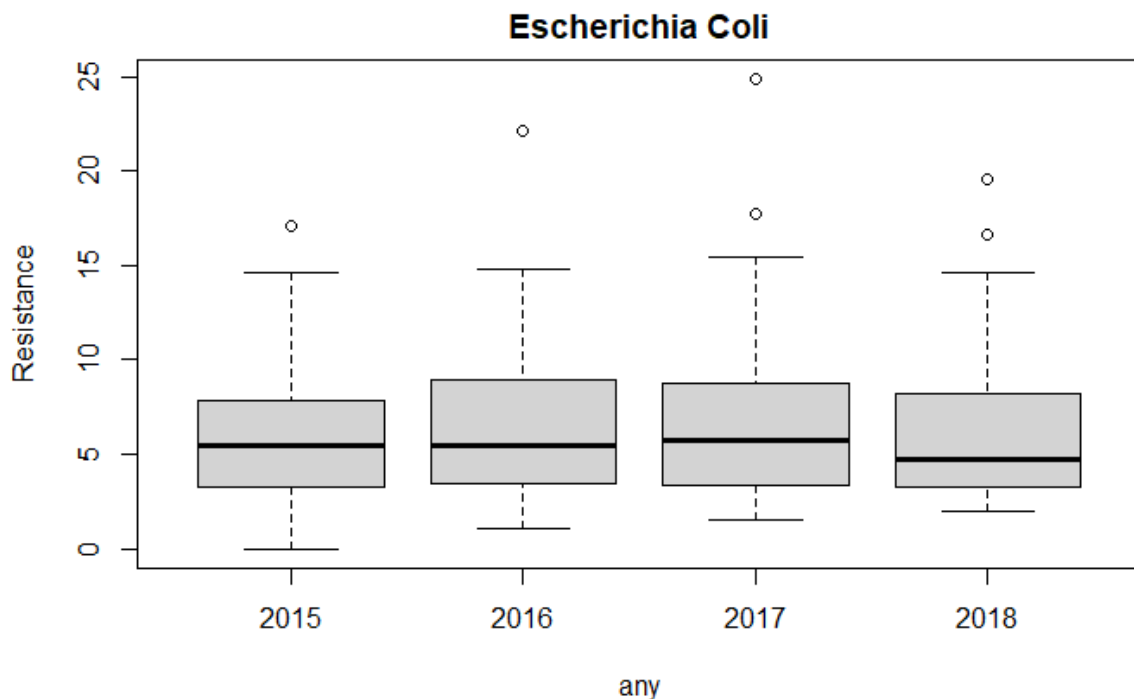


Figura 1: Boxplot de la resistència d'Escheríchia coli als antibiotics estratificat per anys

Pel que es pot veure a simple vista en aquest *boxplot*, fet amb les dades dels 30 països de l'estudi, és que no s'ha vist diferències notables entre les caixes dels anys 2015 fins al 2018. El que s'ha pogut veure més clarament és el fet que l'any 2015 el bacteri ha tingut una alta resistència a Eslovàquia, l'any 2016 a Bulgària i, tant en l'any 2017 com 2018, hi ha hagut una

resistència anormalment alta tant en Eslovàquia com en Bulgària. La resistència de la majoria de països ha estat inferior al 10% mentre que, els 2 països que s'han mencionat anteriorment, ha estat superior al 15% en l'últim any que s'ha mirat.

2.3.2. *Klebsiella pneumoniae*

Klebsiella són bacteris gramnegatius que en alguns casos provoquen infeccions de les vies urinàries o l'aparell respiratori en pacients que es troben en hospitals o en centres sanitaris de llarga estada. Es propaga ràpidament de pacients a personal i el material sanitari a través de les mans i, aquests, ho propaguen a altres pacients per ferides o pels catèters⁸.



Font: shutterstock.com

Alguns dels tipus d'infecció que pot provocar aquest bacteri són infecció de tracte urinari, infecció de tracte respiratori inferior, infecció intraabdominal, i infecció del torrent sanguini.

La *Klebsiella*, igual que l'*Escheríchia coli*, també desenvolupa una resistència als antibiòtics a través dels plasmidis.

⁸ Msdmanuals: infecciones por Klebsiella, Enterobacter i Serratia

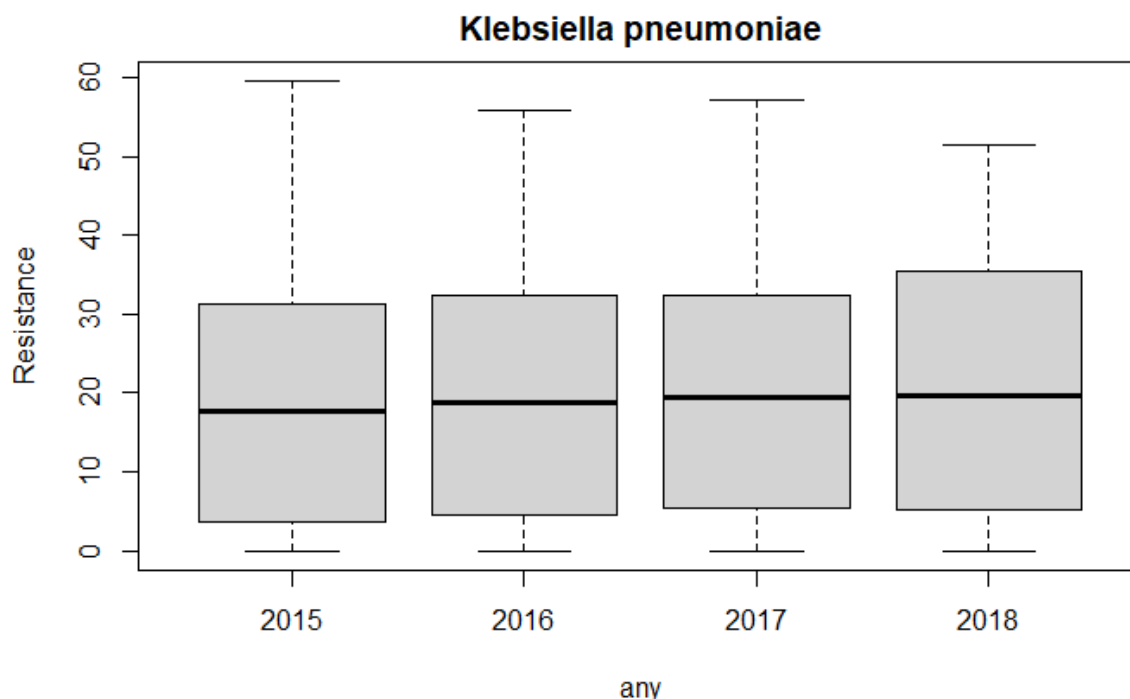


Figura 2: Boxplot de la resistència de *Klebsiella pneumoniae* als antibiòtics estratificat per anys

Es pot veure en aquest *boxplot* que no hi ha diferències molt visibles en el pas dels anys de l'estudi. El més significatiu que s'ha pogut veure és la disminució de la resistència màxima amb el pas dels anys. També es pot veure que hi ha un major rang interquartílic en l'any 2018. En aquest cas es pot veure que no hi ha cap valor extrem.

2.3.3. *Pseudomonas aeruginosa*

La *Pseudomonas aeruginosa* i altres membres d'aquest grup de bacils gramnegatius són patògens oportunistes que sovint causen infeccions intrahospitalàries, especialment en pacients amb assistència respiratòria mecànica, pacients cremats i aquells amb debilitats cròniques. Poden infectar molts llocs, i els quadres solen ser greus. El diagnòstic s'estableix amb el cultiu.

Els bacteris *Pseudomonas* són presents en el sòl i l'aigua en tot el planeta. Aquests bacteris creixen en àrees humides, com ara piques, lavabos, piscines inadequadament clorades i jacuzzis, i en solucions antisèptiques caducades o inactivades. De vegades, aquests bacteris estan presents en les aixelles i àrea genital de les persones sanes⁹.

⁹ Msdmanuals: infecciones por Pseudomonas



Font: Google

Pseudomonas aeruginosa causa pneumònia associada a la salut, infecció del torrent sanguini i infecció del tracte urinari.

Aquest bacteri és intrínsecament resistent a la majoria d'agents microbians per la seva capacitat selectiva per evitar que diverses molècules d'antibiòtics penetren a la membrana externa o expulsar-les si entren a la cèl·lula.

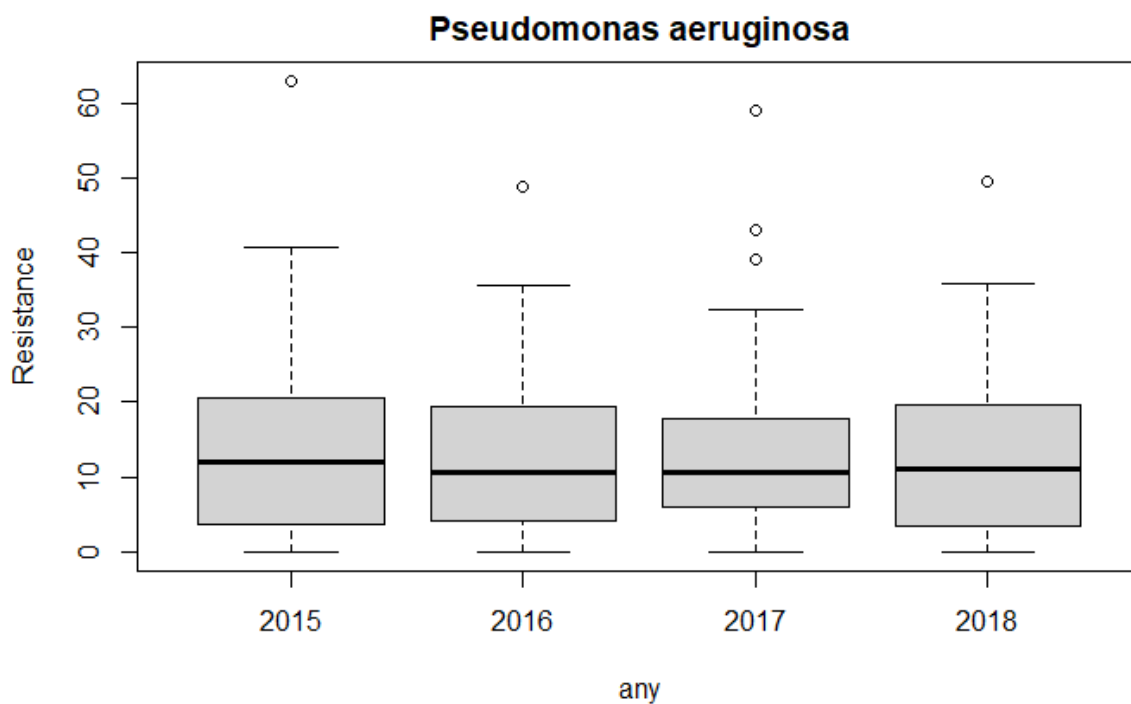


Figura 3: Boxplot de la resistència de *Pseudomonas aeruginosa* als antibiòtics estratificat per anys

En el *boxplot* de la *Pseudomonas aeruginosa* tampoc sembla que hi hagi hagut una gran diferència, tot i que es pot veure que el IQR hi ha disminuït de l'any 2015 al 2017 i, després, ha crescut a l'any 2018 de nou, quedant amb un IQR similar al de la primera caixa. Això pot ser

degut al nombre d'*outliers* que hi ha en la caixa de l'any 2017. Tot i que tant en l'any 2015 com en el 2018 la part superior de la caixa ha estat similar, en l'any 2015 el màxim és superior que el 2018, superant el 40% de resistència.

Romania ha estat *outlier* els 4 anys, en els anys 2015, 2016 i 2018 ha estat l'únic i en l'any 2017 també han sigut *outliers* Letònia i Eslovàquia.

2.3.4. *Acinetobacter* spp.

Les espècies del gènere *Acinetobacter* són microorganismes gramnegatius que poden causar infeccions supuratives en qualsevol sistema d'òrgans; sovint, aquests bacteris són patògens oportunistes en els pacients hospitalitzats. Les espècies més representatives són *Acinetobacter baumannii*, que causa el 80% de les infeccions, i *A. calcoeticus*.



Font: Google

Les espècies de l'*Acinetobacter baumannii* estan associats a infeccions de salut, de torrent sanguini, del tracte urinari i infeccions de ferides. Resulta perillós per a les persones d'edat avançada. Les espècies que pertanyen a l'*Acinetobacter baumannii* són resistents a la majoria d'antibiòtics a causa de la seva capacitat per evitar que diverses molècules puguin penetrar a la membrana externa. La resistència adquirida és deguda a una sèrie de canvis en el cromosoma i l'adquisició de gens resistents a través del plasmidi¹⁰.

¹⁰ Msdmanuals: infecciones por Acinetobacter

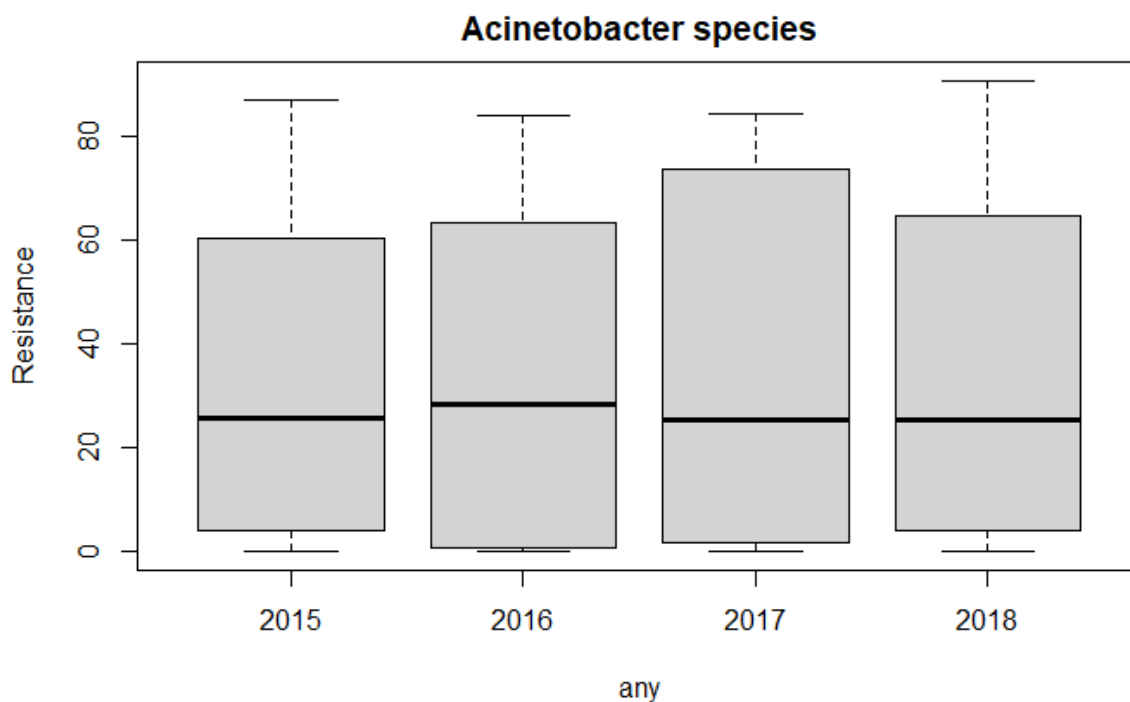


Figura 4: Boxplot de la resistència de *Acinetobacter species* als antibiòtics estratificat per anys

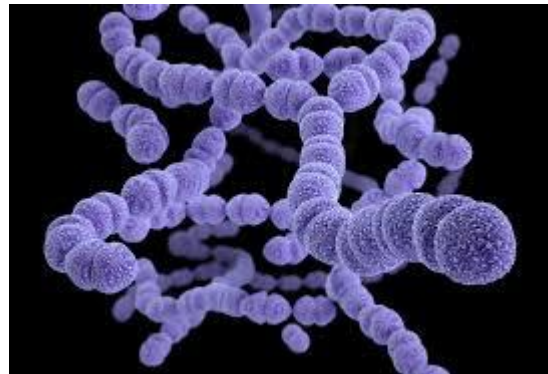
Es pot veure en aquest *boxplot* que aquest bacteri té una resistència molt superior a la resta de bacteris, tenint el límit superior de la caixa amb una resistència superior al 60% i el màxim per sobre del 80% en tot moment. Es pot veure un creixement en el IQR des de l'any 2015 al 2017 i, en l'any 2018, un comportament similar al 2016, sent superior al del 2015 i inferior al del 2018. També es pot veure que no hi ha cap valor extrem.

2.3.5. *Streptococcus pneumoniae*

Bacteris grampositius amb forma d'esfera (cocos) que es poden agrupar formant cadenes. Són patògens gairebé exclusivament humana causant d'un gran nombre d'infeccions (pneumònia, sinusitis, peritonitis, vàlvules cardíques i el torrent sanguini) i de processos invasius severos (meningitis, sèpsia, etc.), particularment en ancians, nens i persones amb el sistema immunològic deprimat. És el principal microorganisme causant de pneumònia adquirida a la comunitat (NAC).

El pneumococ forma part de la microbiota normal humana i es pot trobar a la boca i la faringe, especialment de nens i persones majors de seixanta anys. La transmissió es produeix a través

de gotes de saliva al parlar, tossir, etc. Entra per de les vies respiratòries superiors i de l'oïda mitjana fins a la pneumònia, infecció del torrent sanguini i la meningitis¹¹.



Font: Google

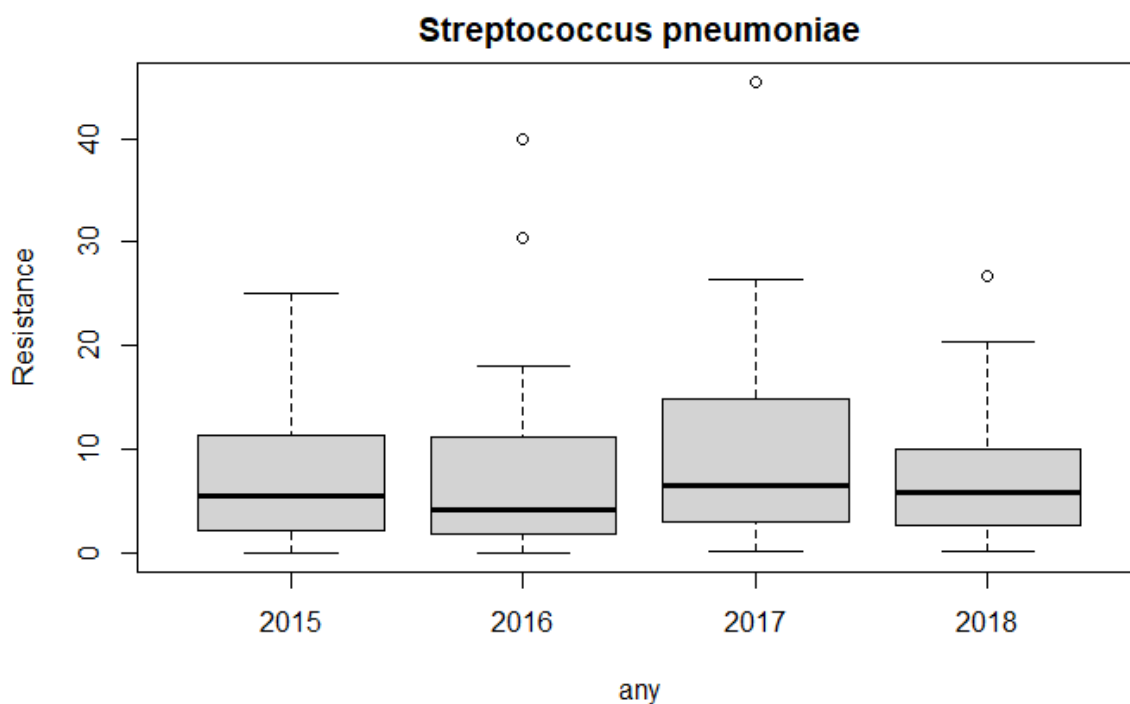


Figura 5: Boxplot de la resistència de *Streptococcus pneumoniae* als antibiòtics estratificat per anys

En aquest *boxplot* es pot veure un creixement en el IQR en l'any 2017 i un decreixement del 2017 al 2018. També es pot veure que en l'any 2015 no hi ha cap *outlier*, mentre que en la resta d'anys sí. En el 2016 hi ha 2 *outliers*, que són Xipre i Romania, en l'any 2017 n'hi ha un, que és Xipre, i en el 2018 n'hi ha un que és Romania.

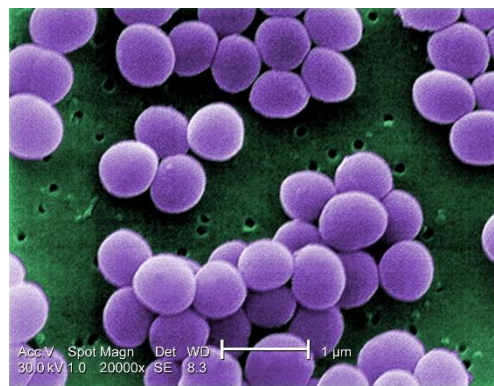
¹¹ Msdmanuals: infecciones por Streptococos

2.3.6. *Staphylococcus aureus*

Staphylococcus aureus és el més perillós de totes els estafilococs, dels quals hi ha molts tipus. Aquests bacteris grampositius en forma d'esfera (cocs) sovint causen infeccions en la pell, però poden causar pneumònia, infeccions de les vàlvules cardíques i infeccions òssies.

Aquests bacteris es contagien per contacte directe amb una persona infectada, a l'utilitzar un objecte contaminat o per inhalació de gotetes infectades que es dispersen a l'esternudar o tossir.

Les infeccions cutànies són freqüents, però el bacteri pot propagar-se a través del torrent sanguini i infectar òrgans distants. Les infeccions cutànies poden causar butllofes, abscessos i enrogiment i inflor de la zona infectada¹².



Font: Google

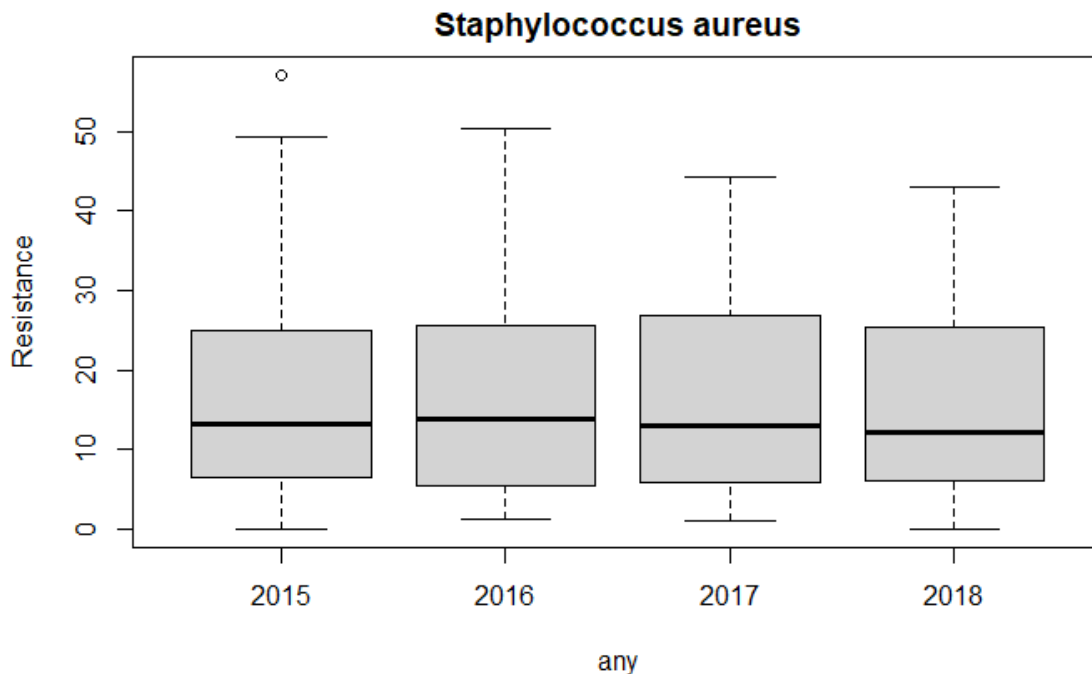


Figura 6: Boxplot de la resistència de *Staphylococcus aureus* als antibiòtics estratificat per anys

¹² Msdmanuals: infecciones por *Staphylococcus*

En aquest *boxplot* es pot veure que no hi ha molta diferència en el IQR, però en el màxim sí que es pot veure que en els dos últims anys ha estat força més alt. Pel que respecta als *outliers*, únicament en tenim un en l'any 2015 i és Romania.

2.3.7. *Enterococci*

Les infeccions enterocóciques estan causades per un grup de bacteris grampositius de la família dels *Streptococcus* denominades enterococs, que resideixen habitualment a l'intestí de persones sanes, però en alguns casos provoquen infecció. Aquests pertanyen a la microbiota bacteriana del tracte gastrointestinal dels éssers humans. Es considera que viu en harmonia amb l'ésser i no causa cap problema, a no ser que entrin a d'altres parts del cos.

Causen infeccions de les vies urinàries (IVU), de les vàlvules cardíaques (endocarditis), de de la pell i dels teixits situats sota la pell (cel·lulitis), de la pròstata (prostatitis) o infeccions de ferides¹³.

Els Enterococs que causen aquestes infeccions són l'*Enterococcus faecalis* i l'*Enterococcus faecium*.



Font: Google

¹³ Msdmanuals: infecciones por enterococos

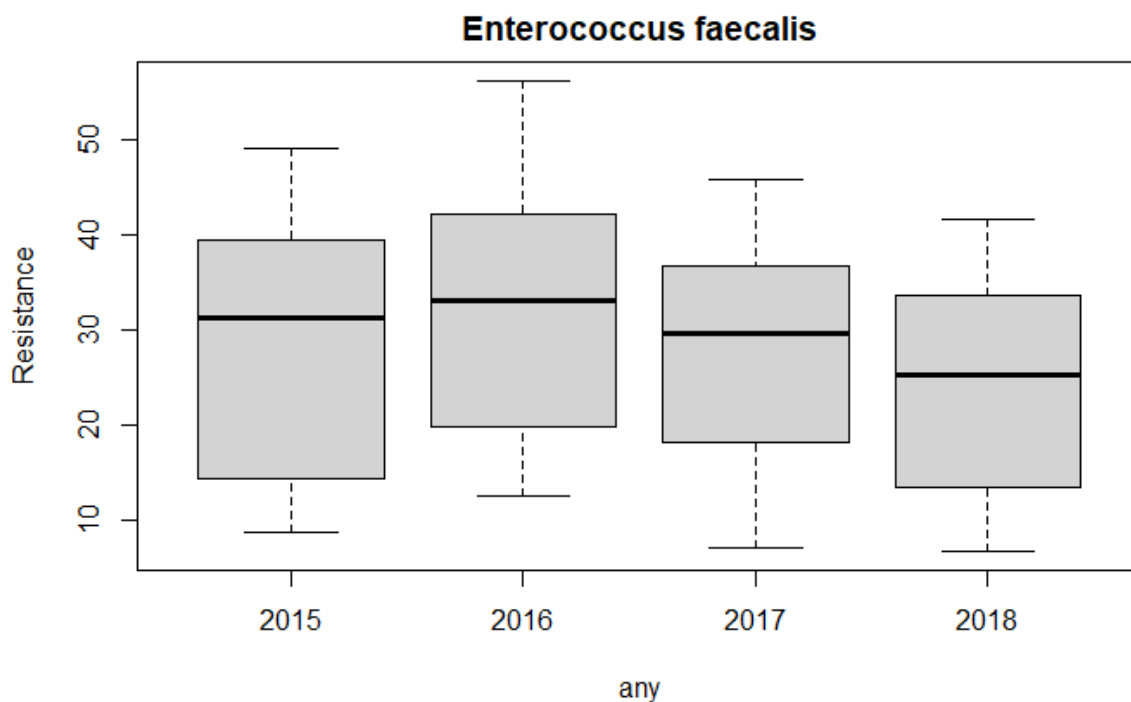


Figura 7: Boxplot de la resistència de *Enterococcus faecalis* als antibiòtics estratificat per anys

En el *boxplot* del *Enterococcus faecalis* es pot veure que des del 2016 fins al 2018 ha decregut la resistència, arribant a aquest últim any a una resistència màxima pròxima al 40%, similar al límit superior de la caixa de l'any 2016. També s'ha vist que no hi ha cap *outlier* en cap any.

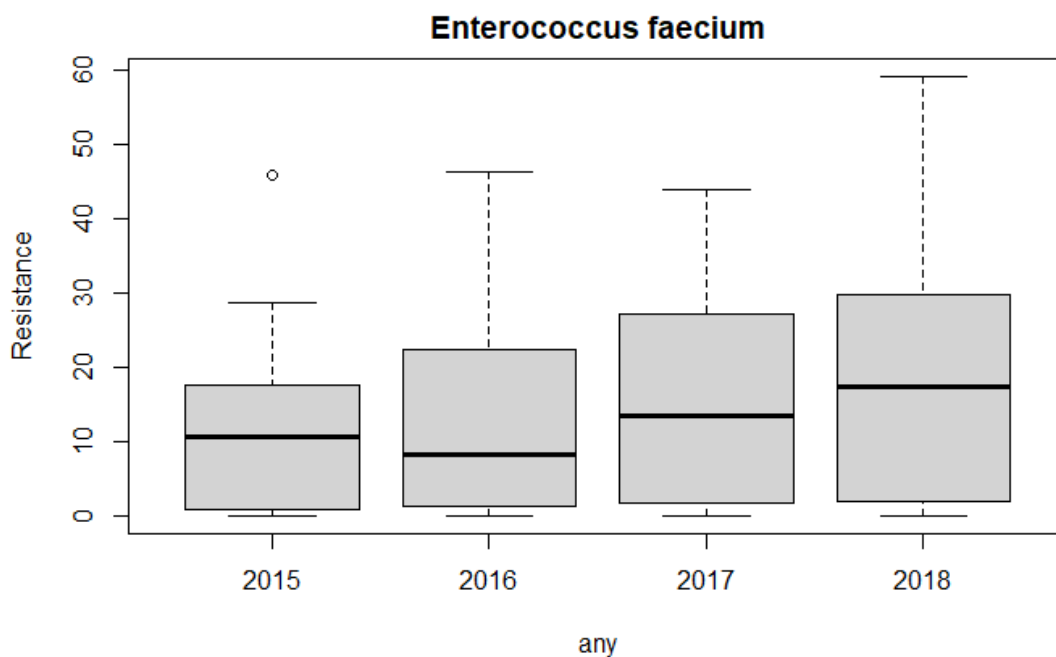


Figura 8: Boxplot de la resistència de *Enterococcus faecium* als antibiòtics estratificat per anys

A diferència del gràfic anterior, aquest sembla que ha incrementat la resistència del bacteri amb el pas dels anys. Cal destacar que el que ha crescut ha estat la part superior, ja que, la part inferior, sembla que no variï amb el pas del temps. També destacar la presència d'un *outlier* en l'any 2015, que fa referència a Romania.

3. Anàlisi de Components Principals

3.1. Què és l'Anàlisi de Components Principals?

L'anàlisi de components principals és una tècnica exploratòria de dades que consisteix en reduir la dimensionalitat de les dades amb la finalitat de mantenir la màxima informació possible amb un volum inferior¹⁴. Aquesta tècnica resulta útil si hi ha una alta correlació entre variables, ja que, amb aquesta tècnica, s'obtenen unes noves variables anomenades components, que ens aporten una informació similar a l'original però, sense ser redundant.

Hi ha dues formes bàsiques d'aplicar el PCA¹⁵:

- Mètode basat en la matriu de correlació, quan les dades no són dimensionalment homogènies o l'ordre de magnitud de les variables aleatòries mesurades no és el mateix.
- Mètode basat en la matriu de covariàncies, que s'utilitza quan les dades són dimensionalment homogènies i presenten valors similars.

S'han analitzat els individus (en aquest treball, països) per respondre diferents qüestions, com per exemple:

- Existeixen similituds entre països per a totes les variables?
- Podem establir diferents perfils de països?

Per altra banda, amb aquest anàlisi també es pot estudiar les relacions lineals de les variables amb diferents objectius com són:

- Veure si es pot resumir la resistència dels països amb un menor nombre de variables.
- Veure si es poden agrupar països per variables.

En aquest treball s'ha realitzat aquest tipus d'anàlisi per a estudiar els individus, els països de la UE i EEE, i veure quins estan pròxims entre si, entenent per proximitat entre països la proximitat entre els resultats d'aquests.

El mètode usat en aquest cas ha estat el mètode basat en la matriu de correlació, mostrada a continuació:

¹⁴ Cienciadatos.net: Análisis de Componentes Principales

¹⁵ rpubs.com: ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

	<i>E. coli</i>	<i>Kleb. pneumo</i>	<i>Aerugin</i>	<i>Acineto</i>	<i>Strep. pneumo</i>	<i>Staphy aureus</i>	<i>Entero. faecalis</i>	<i>Entero. faecium</i>
<i>E. coli</i>	1.0000	0.7107	0.6721	0.6432	0.1498	0.6023	0.5649	0.5481
<i>Kleb. pneumo</i>	0.7107	1.0000	0.7986	0.7774	0.5242	0.6861	0.5091	0.5091
<i>Aerugin</i>	0.6721	0.7986	1.0000	0.6808	0.5288	0.6028	0.6192	0.5563
<i>Acineto</i>	0.6432	0.7774	0.6808	1.0000	0.4390	0.6645	0.4615	0.5650
<i>Strep. pneumo</i>	0.1498	0.5242	0.5288	0.4390	1.0000	0.5347	0.2264	0.3686
<i>Staphy. aureus</i>	0.6023	0.6861	0.6028	0.6645	0.5347	1.0000	0.3351	0.5339
<i>Entero. faecalis</i>	0.5649	0.5091	0.6192	0.4615	0.2264	0.3351	1.0000	0.3643
<i>Entero. faecium</i>	0.5481	0.5891	0.5563	0.5650	0.3686	0.5339	0.3643	1.0000

Taula 5: Matriu de Correlacions any 2018

3.2. Per què l'anàlisi de components principals?

L'objectiu general de l'anàlisi de components principals que s'ha dut a terme en aquest treball és identificar patrons a les nostres dades, de manera que puguem analitzar-les, com ja s'ha dit anteriorment, reduint la dimensió de la base de dades original amb una pèrdua d'informació mínima. Explicat amb altres paraules, l'output que es busca és la projecció de la nostra base de dades original en un subespai més petit, però que mantingui una bona representació de les dades i les descrigui tal com cal. En aquest sentit, més endavant veurem com aconseguir aquesta "bona" representació de les dades mitjançant els valors propis i vectors propis de la nostra matriu de dades.

El PCA aconsegueix aquests objectius transformant les variables inicials en un nou (i més petit) conjunt de variables sense que perdem la informació més rellevant que ens aporten aquestes. Les noves variables o components principals no són més que combinacions lineals de les variables originals. Per exemple, el primer component principal és una combinació lineal de les variables originals que captura la variància màxima de la base dades, determinant la direcció de més variabilitat en les nostres dades n-dimensionals (cap component pot capturar més

variabilitat que el primer, ja que els components estan ordenats per variabilitat explicada de manera descendent).

En el cas de les dades d'aquest estudi, totes les variables han estat considerades actives, ja que en la base original únicament hi ha variables numèriques. Si hi haguera variables qualitatives s'hagueren considerat suplementàries, ja que no poden ser utilitzades en aquest anàlisi i R ens hagués donat una sèrie d'errors, tractant-les com a *dummies* o variables artificials.

3.3. Anàlisi de Components Principals en R

Com s'ha dit anteriorment, l'objectiu general de l'anàlisi de components principals és identificar patrons a les nostres dades. Tenim aquest objectiu perquè una de les finalitats de l'estudi era saber la situació actual en la Unió Europea respecte a la RAM. Com en la base de dades original les dades més recents que es tenen són les de l'any 2018, s'ha treballat amb aquestes dades per a dur a terme el PCA i per saber quina és la situació actual.

A l'hora de dur a terme la reducció de la dimensionalitat, no hi ha un nombre de components ideal a utilitzar. En aquest treball el que s'ha fet és veure la proporció de la variància explicada acumulada que hi ha en cada component i seleccionar un nombre de components, en funció de la variància explicada acumula de cadascun.

És necessari comunicar en aquest moment, que l'Anàlisi de Components principals sempre genera els mateixos components, és a dir, independentment del Software que s'utilitzi, el valor dels *loadings* (coeficients de les components) sempre serà el mateix¹⁶. Tot i tenir sempre el mateix valor, el signe no té per què ser el mateix, ja que el vector de *loadings* determina la direcció del component i aquesta direcció sempre és la mateixa, independentment del signe. Aquests valors es poden veure en la taula 6.

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>
<i>Echer coli</i>	0.07615	0.08508	0.02946
<i>Kleb pneumo</i>	0.36845	0.12079	0.47324
<i>Pseudo aeru</i>	0.22101	0.09193	0.24077

¹⁶ cienciadedatos.net: Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE

<i>Acinetobacter</i>	0.81385	-0.39235	-0.40478
<i>Strep pneumo</i>	0.07309	0.02764	-0.03084
<i>Staphy aureus</i>	0.30738	0.18252	0.62067
<i>Enter faecalis</i>	0.09221	0.23756	-0.19182
<i>Enter faecium</i>	0.19725	0.85162	-0.36106

Taula 6: vectors propis per a els 3 components principals

Una pregunta que es sol fer un cop dut a terme la PCA és quanta informació és capaç de capturar cada component principal que s'ha obtingut. Per a respondre aquesta pregunta es recorre a la proporció de la variància explicada per cada component principal.

Assumint que les variables s'han escalat per a tenir mitjana zero i variància 1, la variància total present en les dades es defineix com¹⁷

$$\sum_{j=1}^p \text{Var}(X_j) = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_{ij}^2$$

On p és el nombre de variables (tipus de bacteris) i n la grandària de la mostra (nombre de països) en el nostre cas.

Mentre que la variància explicada pel component m és

$$\frac{1}{n} \sum_{i=1}^n z_{im}^2 = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{jm} x_{ij} \right)^2$$

Tenint en compte que aquestes components recullen la màxima variabilitat, amb la restricció:

$$\sum_{j=1}^p \phi_{j1}^2 = 1$$

Per tant, la proporció de la variància explicada pel component m ve donada pel rati

¹⁷ cienciadedatos.net: Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE

$$\frac{\sum_{i=1}^n \left(\sum_{j=1}^p \phi_{jm} x_{ij} \right)^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2}$$

Donada una matriu de dades X, amb n observacions i p variables, el nombre possible de components és el min(n-1, p). Quan es duu a terme un PCA per a reduir la dimensionalitat, interessa utilitzar el mínim nombre de components possibles, sempre que expliquin una quantitat d'informació suficient. No existeix un mètode per saber el nombre de components ideal o una resposta correcta si es pregunta quants components utilitzarem. A continuació, en la taula 7 i la figura 9, apareix la proporció de la variància i la proporció de la variància acumulada del PCA fet en aquest treball. S'ha triat un nombre de components que s'ha cregut adequat tenint en compte l'increment de la variància acumulada, és a dir, en el moment en què la variància acumulada ha deixat de créixer de manera substancial, s'ha rebutjat les components que hi havia posteriorment.

<i>Component</i>	<i>Proporció variància</i>	<i>Proporció variància acumulada</i>
<i>PC1</i>	0.7422	0.7422
<i>PC2</i>	0.1358	0.8780
<i>PC3</i>	0.0614	0.9394
<i>PC4</i>	0.0474	0.9868
<i>PC5</i>	0.0086	0.9954
<i>PC6</i>	0.0024	0.9978
<i>PC7</i>	0.0019	0.9997
<i>PC8</i>	0.0003	1.0000

Taula 7: Proporció de la variància per cada component

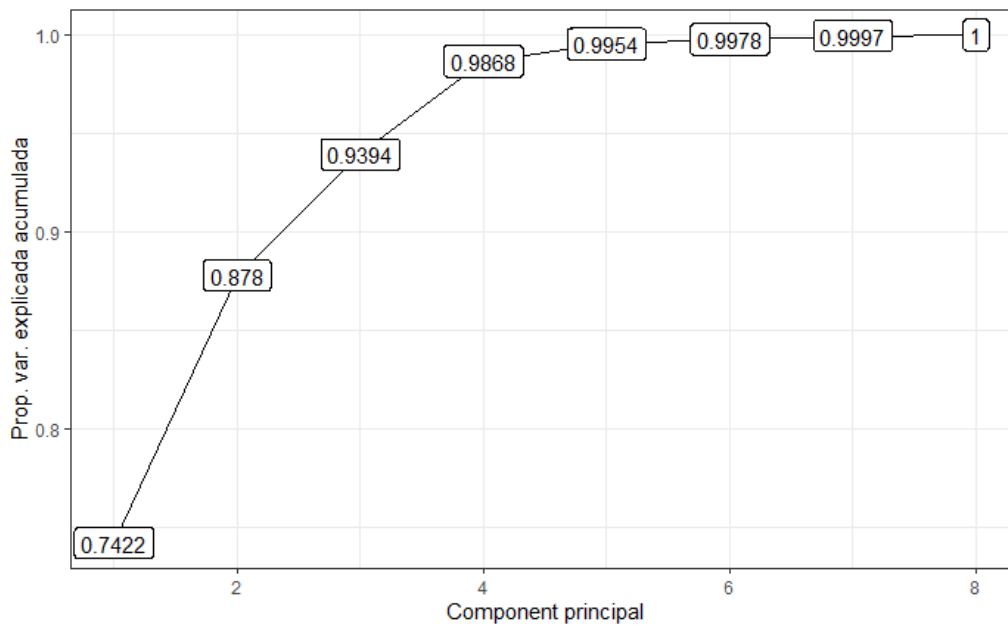


Figura 9: Proporció de la variància explicada acumulada

En la figura i la taula mostrades anteriorment, s'ha vist que el primer component (PC1) explica un 74.22% de la variància observada en les dades. El segon Component principal (PC2) explica un 13.58% i el tercer un 6.141%. S'ha triat treballar amb els 3 primers components principals, ja que han capturat la major part de la informació, és a dir, no es perd res important si es decideix ignorar la resta de components. Aquesta decisió s'ha pres després d'observar la figura 9, on es pot veure que comença amb una gran pendent en els dos primers components. Després, l'increment, té una disminució de manera substancial, de manera que a partir de la quarta component tenen un 2% d'explicació entre tots. Treballant amb els 3 primers components s'ha aconseguit explicar un 93.94% de la variància observada.

Un aspecte clau del PCA és la interpretació dels components, ja que aquesta no ve donada a priori, sinó que serà deduïda després d'observar la relació dels components amb les variables inicials (s'haurà doncs d'estudiar tant el signe com la magnitud de les correlacions)¹⁸. En els següents subapartats del treball s'ha dut a terme l'anàlisi tres cops. El primer cop s'ha treballat amb els dos primers components. Després, s'ha treballat amb el primer i el tercer component i, per acabar, s'ha treballat amb el segon i tercer component. L'anàlisi en tots els casos comença amb una figura on es mostren totes les observacions en un gràfic que té les dues dimensions en les quals es treballa en el subapartat corresponent com a eixos X i Y. Un cop s'analiza el gràfic

¹⁸ cienciadedatos.net: Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE

es busca patrons en funció de les característiques dels països observats, com pot ser la situació geogràfica. Després, s'analitzarà una altra figura, en la que es mostren les projeccions de les diferents variables en un gràfic que té les mateixes dimensions que l'anterior figura com a eixos X i Y. En aquest cas la finalitat de l'anàlisi és observar el vector referent a cada variable respecte les dues dimensions i, també, observar els diferents vectors per veure la correlació que hi ha entre les diferents variables del gràfic.

3.3.1. Components principals 1 vs 2

En aquest apartat s'han comparat el component principal 1 i el component principal 2. Com s'ha vist en ta taula 1, aquestes dues components expliquen un 87.8% del model.

En el gràfic que es pot veure a continuació (figura 10) està representat visualment les puntuacions del segon component principal enfront de les puntuacions del primer. Tot seguit, s'ha fet un *biplot* (figura 11) en el qual es pot observar els coeficients de cada variable per al primer component enfront del segon.

En aquest primer gràfic s'observa la distribució dels països per veure si estan distribuïts de manera aleatòria o segueixen alguna distribució lògica. Per a observar això, s'han dividit els països en 4 zones, que són les següents:

- Europa Nord o Septentrional: Zona d'Europa formada per Dinamarca, Estònia, Finlàndia, Irlanda, Islàndia, Letònia, Lituània, Noruega, Països baixos, Regne Unit i Suècia.
- Europa Meridional o Sud: Zona d'Europa formada per Croàcia, Eslovènia, Espanya, Grècia, Itàlia, Malta i Portugal.
- Europa Est o Oriental: Zona d'Europa formada per Bulgària, Eslovàquia, Hongria, Polònia, República Txeca i Romania.
- Europa Occidental: Zona d'Europa formada per Àustria, Bèlgica, França, Luxemburg i Països baixos (situats tant a la zona nord com a l'occidental per situació geogràfica).

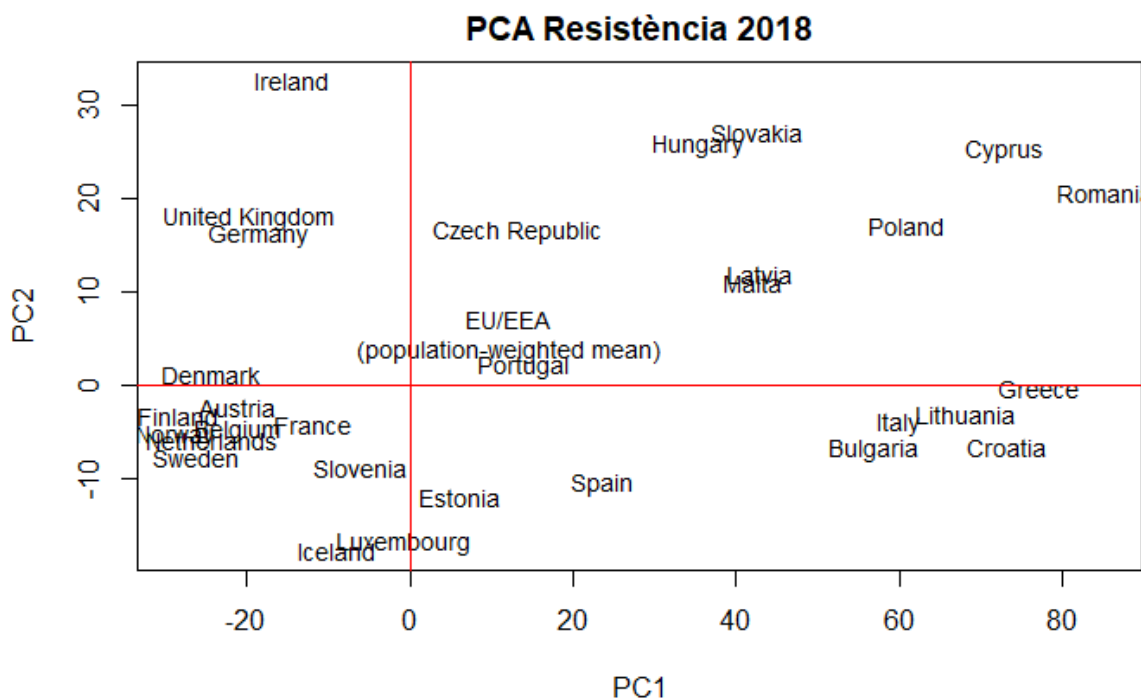


Figura 10: Puntuacions del PC1 enfront PC2

En el gràfic que s'ha presentat anteriorment, es pot veure una sèrie de conglomerats, on el principal està format per Dinamarca Àustria, Finlàndia, Bèlgica, Noruega Països baixos, Suècia i França. Hi ha un altre conglomerat format per Grècia, Lituània, Itàlia, Bulgària i Croàcia. No s'ha trobat cap altre conglomerat, però sí que s'ha vist parelles de països situats molt prop entre si, com ha estat el cas de Malta amb Letònia o Regne Unit amb Alemanya.

S'ha vist que en la distribució dels països en aquest gràfic hi ha una relació entre la zona a la qual pertany el país amb algun component. Un clar exemple d'això s'ha vist en els països d'Europa Nord, que no ha demostrat cap patró de comportament observable en el PC2, però, en canvi en el PC1, s'ha vist que tots se situen a la mateixa altura menys Estònia, situada lleugerament més a la dreta, Letònia i Lituània, situats a l'altra part del gràfic. El mateix ha passat amb els països d'Europa Occidental, que segueixen el mateix comportament que Europa Nord en PC1. Tant és així que el conglomerat més gran del gràfic, del qual s'ha parlat anteriorment, està format per països d'aquestes dues zones.

Mentre que els països d'Europa occidental i septentrional estan a l'esquerra del gràfic, els països de la zona d'Europa de l'est se situen a la part dreta, on sembla que també segueixen un comportament similar al de les anteriors zones, tot i que, en aquest cas, els països tenen una

separació superior entre si. També s'ha vist que, dintre dels països d'aquesta zona, destaca República Txeca, estant situat al centre del mapa i no a la dreta com la resta.

Mentre que en totes les zones observades anteriorment s'ha vist una relació en el PC1, en aquest cas, es pot veure com, en la zona d'Europa Meridional gairebé tots els països tenen una puntuació similar, situada entre el -10 i el 0 de PC2, menys Portugal, situat lleugerament per sobre del 0, i Malta, situat a una altura considerablement superior a la resta de països de la taula.

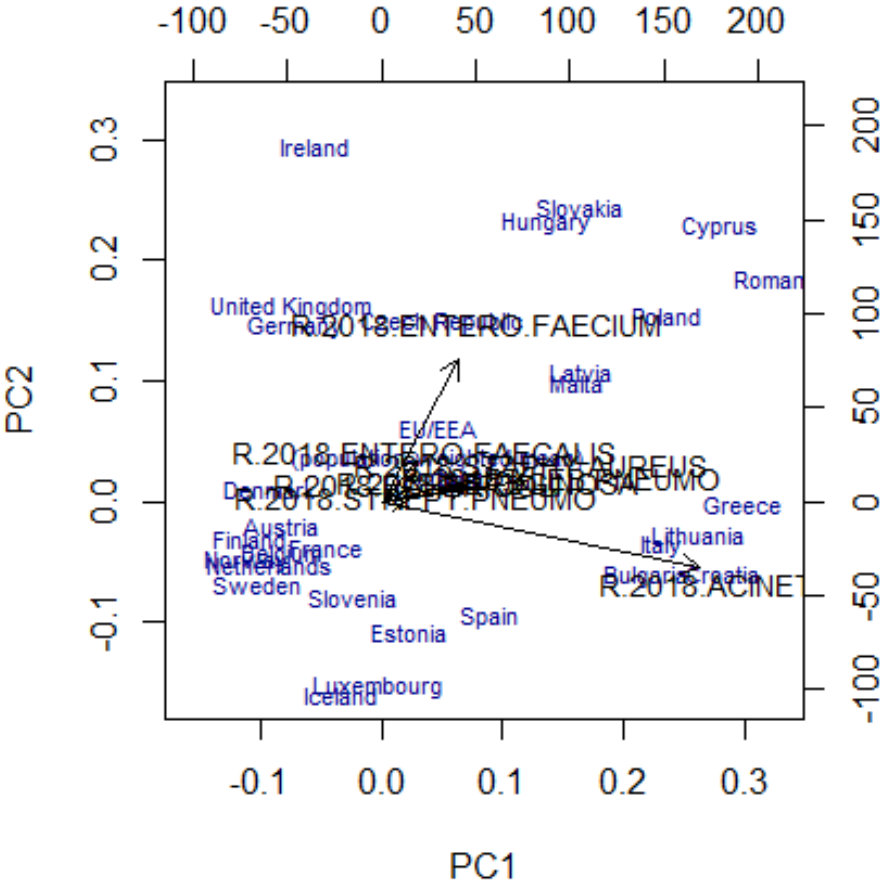


Figura 11: Projeccions de les variables per a PC1 i PC2

En aquest gràfic s'ha pogut observar la influència que ha tingut cada variable en cada component. Les influències poden anar de -1 a 1, en aquest cas l'escala ha estat menor, ja que no tenen una influència suficientment forta perquè el gràfic estigui a escala completa. Aquest valor depèn de com afecta la variable al component, com més influència tingui la variable en el component, més lluny estarà del valor 0 i, com més prop estigui del 0, per tant, menys influència té la variable.

Podem veure que per a la PC1 totes les influències han estat positives, ja que no hi ha cap fletxa que hi hagi anat cap a l'esquerra. Tot i ser totes positives, es veu a simple vista que la majoria

de variables no tenen gaire influència, ja que estan pròximes a l'eix 0 per al PC1. Així, s'ha pogut veure que PC1 es veu caracteritzat principalment per la variable que fa referència al bacteri *Acinetobacter*.

Pel que fa a la PC2, s'ha observat que no hi ha una influència tan forta per part de cap variable com en l'anterior cas. Tenint en compte que totes les variables estan situades en un interval d'entre -0.1 i 0.2, que no hi ha cap variable que faci referència al PC2, tot i que la variable *Enterococcus faecium* té una influència superior a les altres.

També cal destacar la gran correlació que s'ha mostrat entre les variables, sobretot entre *Escheríchia coli*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Straphylococcus aureus* i *Enterococcus faecalis*. Això s'ha pogut veure a l'hora d'observar els angles entre vectors, ja que els angles formats per les fletxes és petit i, com més petit és, més correlacionades estan les variables per aquestes dimensions.

3.3.2. Components principals 1 vs 3

En aquest cas, en compte de comparar en compte de comparar PC1 amb PC2, s'han comparat el PC1 amb PC3. Com la proporció de la variància del PC1 és del 74.22% i la del PC3 és del 6.141%, La proporció de la variància acumulada per aquests dos components és del 80.16% del model.

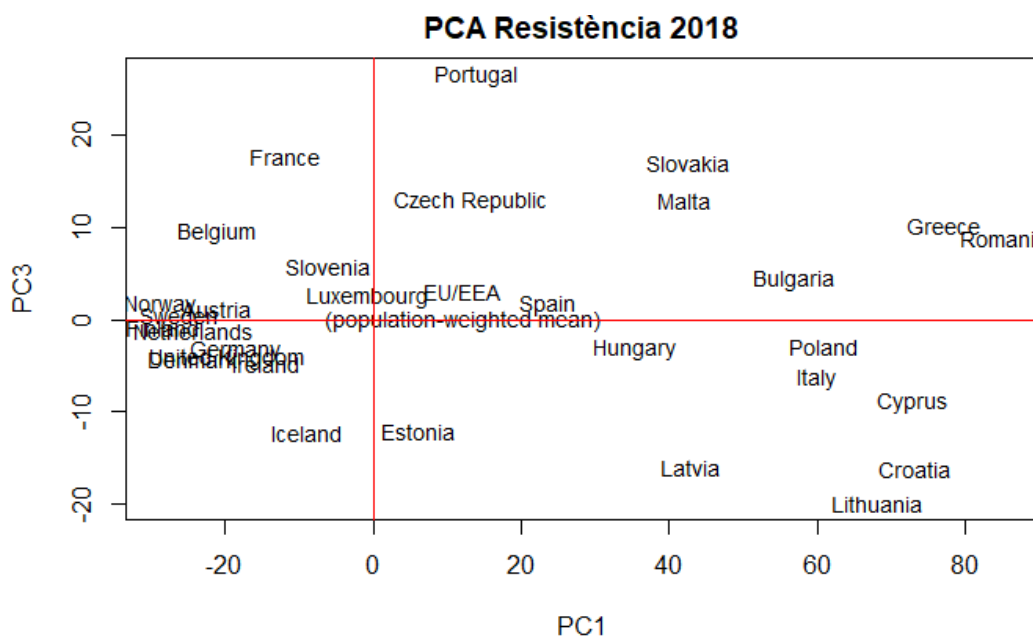


Figura 12: Puntuacions del PC1 enfront PC3

En la figura 12 s'ha dispersat un dels dos conglomerats que havíem vist anteriorment. Això es deu a la substitució del PC2 pel PC3. L'altre conglomerat segueix estant, tot i que ha canviat lleugerament els països del qual està compost.

Anteriorment, en la figura 10, ja s'ha comentat el comportament de la zona d'Europa Nord en el PC1, on s'ha dit que normalment els països d'aquesta zona estan situats a la part esquerra del mapa, ja que tenen una puntuació negativa. En aquest cas, a diferència de l'anterior, tenen una tendència en els dos components, és a dir, en PC1 es manté el comportament de l'altre gràfic, però en PC3 també tenen un comportament no aleatori. S'afirma això, ja que s'ha vist que totes les observacions que fan referència a aquesta zona tenen un valor negatiu en el PC3, menys Noruega i Suècia que estan lleugerament per sobre de la línia vermella que representa el valor 0. Per aquest motiu estan totes les observacions de la zona nord d'Europa a una distància relativament petita en aquest gràfic. Pel que fa a la zona d'Europa sud no hi ha nova informació, ja que els diferents països estan dispersats en tot el mapa i no sembla que hi hagi cap agrupament d'observacions per aquests dos components. Tampoc hi ha informació nova en la zona d'Europa de l'est, ja que estan dispersats a causa del PC3. Pel que fa a la zona d'Europa Occidental, s'ha vist que hi ha dispersió en el PC3, però tot i això segueix havent un conglomerat, ja que estan relacionats pel PC1. Cal recordar que aquest conglomerat no està format únicament per països d'Europa occidental, sinó que també hi ha països d'Europa del Nord.

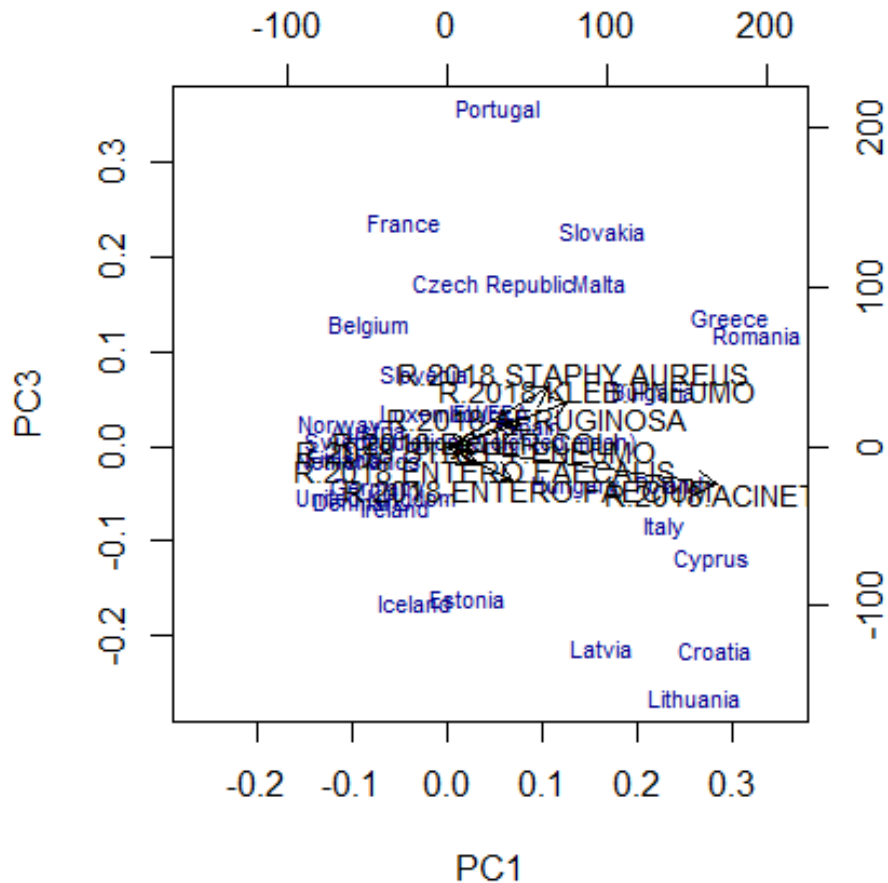


Figura 13: Projeccions de les variables per a PC1 i PC3

En la figura 13 s'ha vist la influència que ha tingut les variables en cada component. La influència de les diferents variables sobre PC1 ja s'ha vist anteriorment, així que en aquest cas s'ha observat la influència sobre PC3.

En aquest cas, sembla que les variables no tenen gaire influència, ja que, totes les variables es troben dintre de l'interval -0.1 i 0.1. Dintre d'aquest interval, la variable *Staphylococcus aureus* resulta la més influent, tenint una influència positiva superior a les altres, tot i que no resulta gaire més influent que altres variables negatives, com pot ser l'*Acinetobacter*.

Mentre que en PC1 totes les variables resulten positives, ja que totes les fletxes estan cap a la dreta, en PC3 n'hi ha 4 de positives i 4 de negatives, fàcilment diferenciables, ja que l'extrem de les fletxes de les variables positives estan per damunt del seu origen, és a dir, que tenen una pendent positiva, mentre que les variables amb influència negativa tenen les fletxes amb pendent negativa.

Pel que fa a l'angle de les fletxes segueix sent baix. Això ens indica que segueix havent-hi una correlació forta entre variables tot i haver canviat un component.

3.3.3. Components principals 2 vs 3

Per acabar la comparació de components principals s'han comparat els components PC2 i PC3. En aquest cas, la proporció de la variància acumulada que ha estat molt inferior als altres casos, això es deu al fet que en aquest cas s'ha ignorat PC1, component que explicava un 74.22% de la proporció de la variància. En aquest cas la proporció de la variància acumulada ha estat del 19.81%.

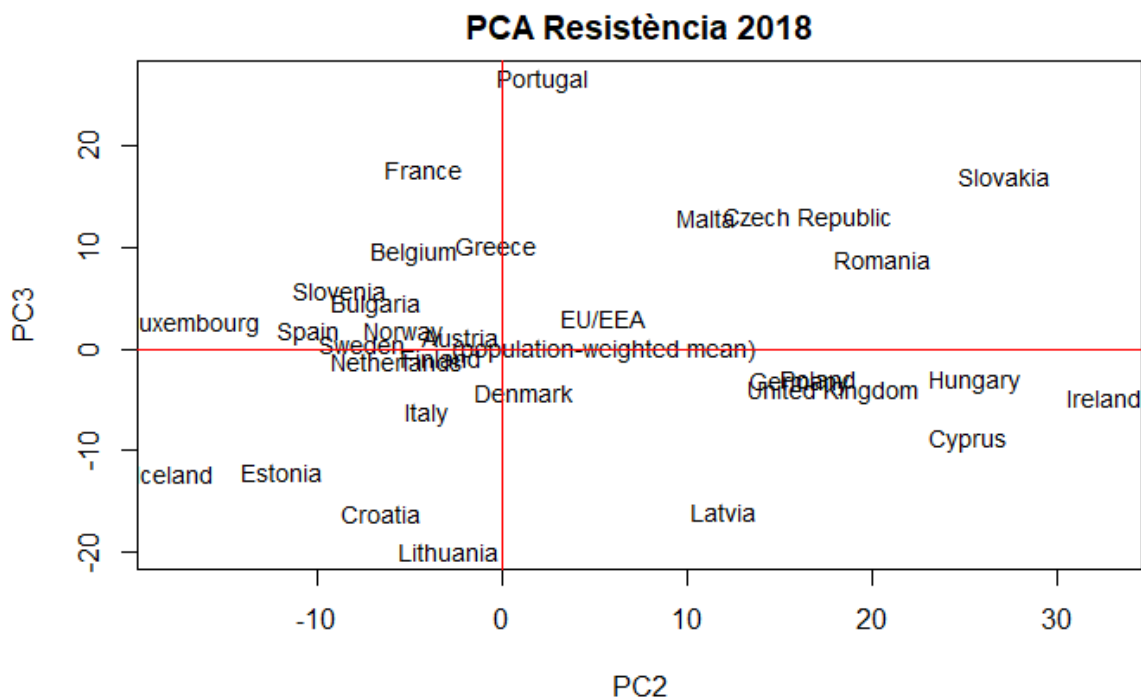


Figura 14: Puntuacions del PC2 enfront PC3

En la figura 14 es pot veure un parell de conglomerats. El més gran està format per Eslovènia, Bulgària, Espanya, Noruega, Suècia, Àustria, Finlàndia i Països Baixos. Tenint en compte que en aquest conglomerat hi ha països de les 4 zones europees, no sembla que hi hagi una aglomeració a causa de la zona geogràfica. Igual que en aquest cas ja esmentat, el mateix passa amb un petit conglomerat format per 3 països de 3 zones geogràfiques europees, que són Polònia, Regne Unit i Alemanya.

Tot i que no hi ha aglomeracions d'observacions per zona geogràfica segueix havent-hi les relacions entre zones i components que s'ha mencionat en els anteriors apartats. Aquests són la zona d'Europa meridional amb PC2, situada la major part dels països lleugerament a l'esquerra del zero, és a dir, amb un valor negatiu. També s'havia comentat anteriorment la relació entre PC3 i els països de la zona Nord d'Europa, on quasi tots estan situats per sota del 0.

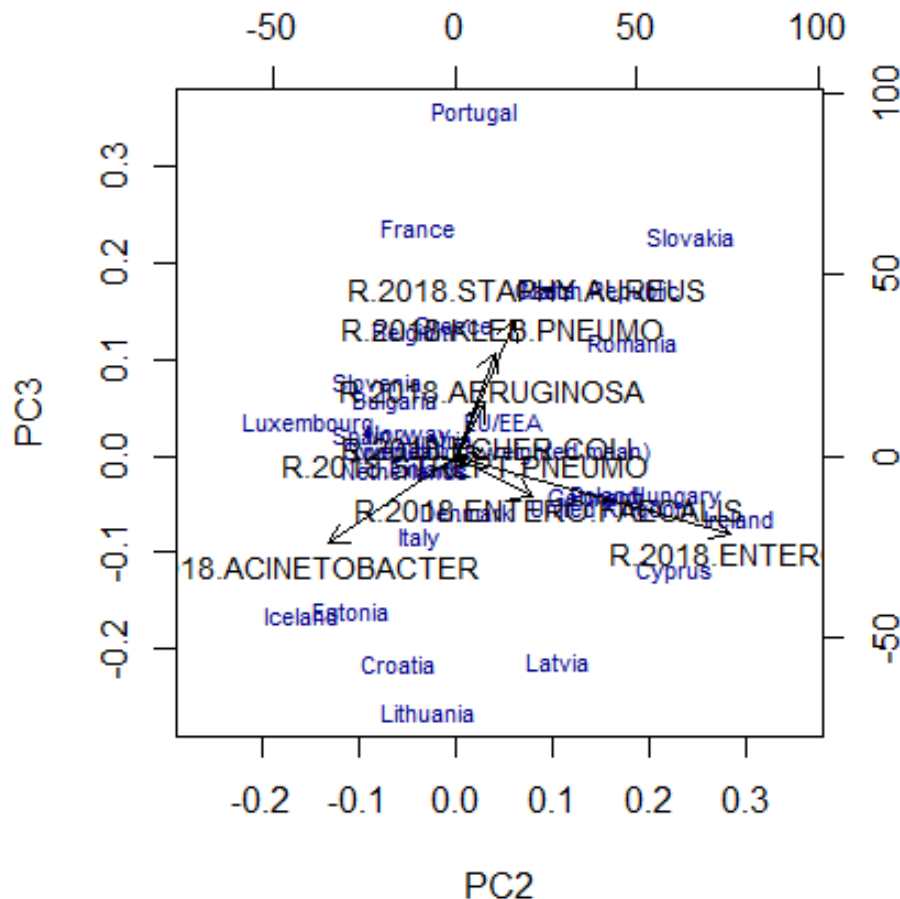


Figura 15: Projeccions de les variables per a PC2 i PC3

En aquest gràfic s'ha vist la influència de les variables sobre PC2 i PC3. E aquest cas es pot observar, a diferència dels apartats anteriors, una variable que és negativa per als dos components, com és el cas del *Acinetobacter*. Això resultava previsible, ja que en els dos gràfics que s'han vist anteriorment, l'únic component on aquesta variable tenia un valor positiu era PC1.

Pel que fa a PC2, es pot veure que té una gran influència la variable *Enterococcus faecium*, molt superior a la resta. En el cas de PC3 en aquest cas, la variable que ha resultat més influent és *Staphylococcus aureus* amb una influència pròxima al 0.15, que, tot i ser superior a totes les altres, segueix sent baixa.

En aquest cas s'ha vist una correlació molt forta de les variables *Staphylococcus aureus*, *Klebsiella pneumoniae* i *Pseudomonas aeruginosa* entre si, com també s'ha vist una forta correlació entre les dues variables d'*Enterococcus*. Aquests dos grups de variables no sembla que tinguin cap relació entre si, ja que formen un angle pròxim als 90 graus. Una altra correlació forta que hi ha és la que hi ha entre les variables *Acinetobacter* i *Escherichia coli*, però, en aquest cas, a diferència dels altres casos, és una correlació negativa. Es veu en el gràfic la correlació negativa que tenen, ja que l'angle que formen és pròxim als 180 graus. Això ens indica que en el moment que incrementin els valors en una d'aquestes variables, decreixerà els valors de l'altra variable, és a dir, tenen una proporcionalitat inversa.

3.4. Interpretació del PCA

En l'última part d'aquest bloc s'interpreten els resultats obtinguts en els PCAs. Per a facilitar la interpretació s'ha introduït la taula que es pot veure a continuació, ja que, en proporcionar-nos valors numèrics, es pot contrastar la informació que s'ha proporcionat en els gràfics anteriors amb una millor precisió.

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>
<i>Echer coli</i>	0.71526	0.40082	0.08024
<i>Kleb pneumo</i>	0.87267	0.38060	0.23157
<i>Pseudo aeru</i>	0.76903	0.43546	0.19361
<i>Acinetobacter</i>	0.97370	0.05081	-0.32154
<i>Strep pneumo</i>	0.52637	0.26086	0.18117
<i>Staphy aureus</i>	0.75701	0.38146	0.36699
<i>Enter faecalis</i>	0.52754	0.36770	-0.08187
<i>Enter faecium</i>	0.66647	0.82746	-0.24612

Taula 7: Matriu de components per a les 3 components principals

En la taula 7 es pot veure la matriu de components, matriu que es caracteritza per mostrar les correlacions entre cada variable i cada un dels 3 components principals.

Aquesta taula ens mostra que en el primer component principal tots els valors són positius. Això ja es podia saber interpretant els gràfics on totes les fletxes apuntaven a la dreta per al PC1. També s'havia vist anteriorment, per a PC1, que la variable amb una correlació més alta ha estat la variable *Acinetobacter*, tenint un valor molt superior a les altres variables per aquest component. En el cas de PC2 s'havia dit que la variable més influent havia estat *Enterococcus faecium* i, tot i ser cert, no s'havia apreciat una diferència tan gran respecte a les altres variables. Pel que fa al 3r component, es pot veure que la variable més correlacionada ha estat *Staphylococcus aureus*, seguida de la *Klebsiella pneumoniae*.

Quant a les observacions, es pot veure que els països situats en la zona nord d'Europa tenen una baixa RAM per part del bacteri *Acinetobacter*, representat per PC1. El mateix passa amb el bacteri *Staphylococcus aureus*, representat pel component principal PC3. En general no presenta resistències gaire altes, menys en *Enterococcus faecium*, representat per PC2.

Pel que fa als països d'Europa sud, tot i que no segueixen un comportament igual en PC1, sol tenir un valor més elevat que els de la zona nord. Això es pot veure reflectit en la resistència del bacteri *Acinetobacter*, que presenta valors superiors als de l'anterior zona. En canvi, en la variable *Enterococcus faecium* té un valor inferior al de l'anterior zona, a causa d'un valor inferior en PC2. En PC3, igual que en PC1, no s'ha vist un comportament clar en els gràfics anteriors, però sí que estaven en valors alts en comparació a la zona Nord, motiu pel qual el bacteri *Staphylococcus aureus* també presenta una resistència superior.

La zona est d'Europa es caracteritza per tenir valors molt elevats en la PC1. Això explica el perquè dels valors de la variable *Acinetobacter*, que resulta extremadament alta en aquests països. El cas de PC2 resulta similar al de PC1, per això també tenen un valor alt en l'*Enterococcus faecium*. En el cas de PC3, els països estan dispersats, així que no es dirà molt al respecte.

A la quarta i última zona, l'occidental, s'ha observat un comportament similar a la zona nord per a PC1. Es pot observar que els valors de la variable *Acinetobacter* són molt baixos. En el cas de PC2, majoritàriament segueix tenint valors molt baixos, igual que en PC3.

Actualment sembla que les zones d'Europa amb més bona situació han estat la zona nord i la zona d'occident, ja que són les zones amb menys RAM. Cal comentar que dintre de la zona nord, Letònia i Lituània semblen tenir uns nivells de resistència extremadament alts en comparació als altres països de la mateixa zona.

La zona d'Europa Est, en canvi, ha resultat tenir uns valors molt superiors a les anteriors dues zones, igual que la zona sud. Això es podia interpretar d'un principi, on aquestes dues zones mostraven uns valors superiors en PC1, component que mostrava un pes molt superior a la resta de components.

4. Anàlisi Factorial Múltiple

L'Anàlisi factorial múltiple (AFM) és una tècnica estadística per a dades tant numèriques com categòriques¹⁹. Aquest procediment és utilitzat per estudiar la similitud entre individus respecte al conjunt de variables i la relació entre aquestes. Per aplicar aquesta tècnica, s'han de dividir les variables de la base de dades en grups per similituds entre significat de variables, aquests grups han d'estar formats per variables numèriques o categòriques, no poden ser mixtos. Els objectius d'aquest anàlisi són: estudiar les similituds i diferències entre grups i comparar les característiques dels individus des d'un anàlisi per separat.

La metodologia del MFA es divideix en dues fases²⁰:

- S'executa per a cada bloc o conjunt de dades un PCA o un MCA, depenent del tipus de variables que es tenen en cada bloc. S'emmagatzema el valor del primer valor propi de cada anàlisi per a ponderar els diferents blocs més avant, en la segona part de l'anàlisi.
- Es duu a terme un PCA sobre les columnes de totes les taules de manera que cada variable indicador té un pes que és una funció de la freqüència de la categoria corresponent. La ponderació de taules fa possible preveure que els blocs que inclouen més variables no influeixin en l'anàlisi, tot i que en aquest cas, no ens afecta, ja que tots els blocs tenen el mateix nombre de variables.

L'anàlisi Factorial múltiple es pot utilitzar en diversos camps i aquests són²¹:

- Anàlisi sensorial: Un individu és un producte alimentari. Un conjunt de variables inclou variables sensorials, com poden ser la dolçor, amargor, etc. Després hi ha un segon conjunt que inclou variables químiques, com poden ser pH, taxa de glucosa, etc.
- Ecològica: Un individu és un lloc d'observació. Un primer conjunt de variables descriu les característiques del sol; el segon descriu la flora.
- Sèries temporals: S'observen diferents individus en dates diferents. Normalment hi ha dues maneres de definir grups de variables: generalment s'agrupen el conjunt de variables observades en el mateix temps.

¹⁹ factominer.free.fr: Multiple Factor Analysis

²⁰ help.xlstat.com: ANÁLISIS FACTORIAL MÚLTIPLE (MFA) EN EXCEL

²¹ sthda.com: MFA - Multiple Factor Analysis in R: Essentials

En aquest treball el conjunt de variables és de sèries temporals, ja que en aquest cas l'objectiu de l'estudi ha estat estudiar l'evolució de la resistència als antibiòtics, a la Unió Europea, durant els últims anys.

4.1. Anàlisi Factorial Múltiple en R

En la següent part del treball es busca complir l'objectiu de veure l'evolució de la RAM a través dels anys. Per això, en aquesta part del treball, s'ha tingut en compte totes les dades que hi ha a la base, a diferència de la part del PCA, que s'havia utilitzat únicament les dades referents als 2018.

La funció d'R que s'ha fet servir per calcular aquest anàlisi és MFA, funció en la qual s'ha indicat (X), primerament, quina és la base de dades, després, un vector de la mida de cada grup (group), un vector del tipus de variables que hi ha en un grup (type), un vector dels noms dels grups suplementaris (num.group.sup) que en el nostre cas és NULL, ja que no hi ha grups suplementaris, el número de dimensions a mantenir en els resultats (ncp) i un vector amb els noms de cada grup (name.group). En el cas d'aquest treball s'han format quatre grups de variables numèriques. Aquests grups estan formats pels diferents bacteris que s'han estudiat.

Els quatre grups de variables creats fan referència a l'any de les dades, que són 2015 per al primer bloc, 2016 per al segon, 2017 per al tercer i 2018 per al quart.

A continuació es mostren els subapartats en el que està dividit el MFA. En el primer subapartat es mostra el percentatge de variància total explicada per les dimensions creades per a saber com són d'informatives les primeres dimensions, les 7 primeres en el cas de la taula 8 i les 10 primeres en el cas de la figura 16. Un cop s'ha observat la figura i la taula es tria el número de dimensions amb el qual es vol treballar. En els següents apartats s'analitza la contribució de cada grup en les dues dimensions per veure si hi ha algun grup que tingui una contribució major i per veure si hi ha diferències entre anys. Cal recordar que un dels principals objectius del treball ha estat observar si hi ha diferències entre els diferents anys.

El següent subapartat tracta d'observar la contribució dels diferents països per a cada dimensió a través del CTR i el \cos^2 , explicats més avant. Un cop observats els diferents països s'observa l'evolució d'alguns d'aquests amb el pas del temps. En el tercer subapartat d'anàlisi de la contribució en les diferents dimensions s'observa la contribució de les diferents variables a les diferents dimensions. Primerament s'observa la contribució de cada variable en la primera dimensió per veure si predominen les variables d'algun any o bacteri o grup en concret. Posteriorment es duu a terme el mateix anàlisi per a la segona dimensió i per a les dues

dimensions conjuntes. Després s'han observat les correlacions entre les variables i les dimensions, ja que la correlació ens indica quines variables són més importants per explicar la variabilitat del conjunt de dades. Finalment s'analitzarà la relació entre individus i dimensions primer i entre variables i dimensions tenint en compte el quadrat del cosinus. El valor del quadrat del cosinus ens indica com de ben representat un individu o una variable en les dimensions.

4.1.1. Explicació de la variabilitat en cada dimensió

En aquest cas s'ha dut a terme el MFA amb la intenció de reduir la dimensionalitat de les dades, ja que en aquest cas la base consta de 32 dimensions, formada per 8 bacteris dels quals tenim 4 dimensions per cadascun. S'ha observat la proporció de la variància que té cada dimensió per poder seleccionar un nombre de dimensions en el qual es pugui treballar de manera eficient.

	<i>Dim. 1</i>	<i>Dim. 2</i>	<i>Dim. 3</i>	<i>Dim. 4</i>	<i>Dim. 5</i>	<i>Dim. 6</i>	<i>Dim. 7</i>
<i>Var.</i>	3.935	0.590	0.386	0.266	0.190	0.092	0.071
<i>% de Var.</i>	68.679	10.290	6.743	4.641	3.320	1.601	1.247
<i>Acum %</i>	68.679	78.969	85.712	90.354	93.674	95.275	96.522

Taula 8 : Proporció de la variància de les dimensions amb més importància

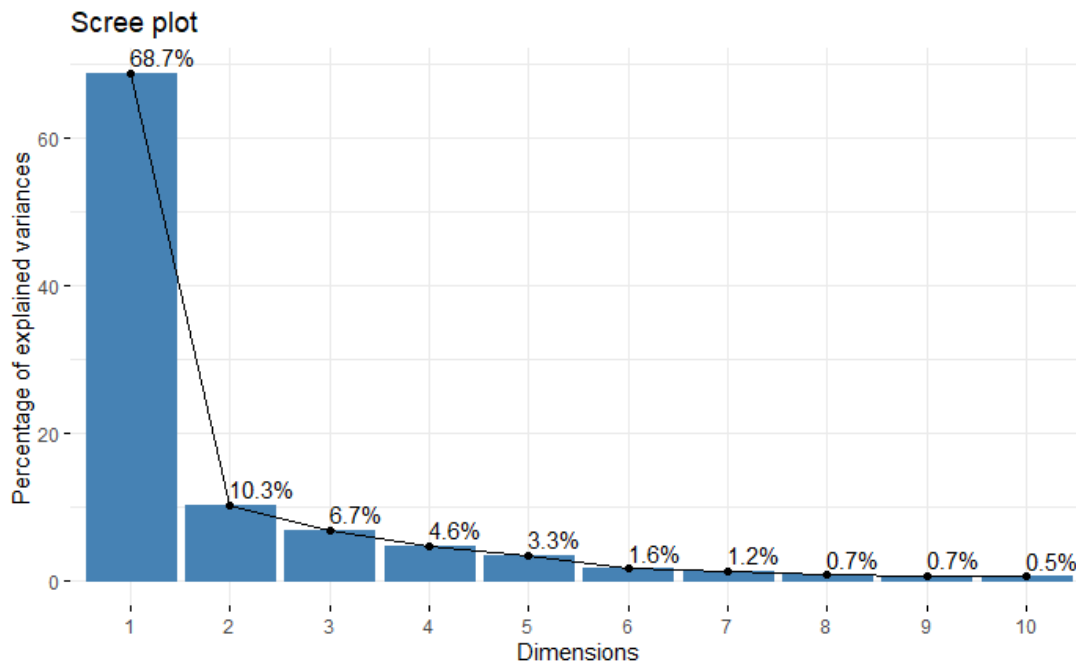


Figura 16: Percentatge de la variabilitat explicat per a cada dimensió del MFA

S'ha vist que les dues primeres dimensions expliquen quasi el 80% de la variabilitat, explicant el 68.7% la primera i un 10.3% la segona. És pots veure en la figura 16 que després del segon factor les variables ja expliquen una un percentatge de la variància bastant baix. Encara que no es pugui veure ni en el gràfic ni en la taula, el nombre de dimensions són 32 i només apareixen les més grans, per aquest motiu el percentatge acumulat de la variància no arriba al 100%.

S'ha triat treballar amb 2 dimensions, ja que, d'aquesta manera, es fa una gran disminució de la dimensionalitat de les dades i ens queda una explicació de la variabilitat acceptable, és a dir, no es perd una quantitat d'informació important.

4.1.2. Contribució de cada grup a les dimensions

<i>Any</i>	<i>Dim 1</i>	<i>CTR</i>	<i>Cos²</i>	<i>Dim 2</i>	<i>CTR</i>	<i>Cos²</i>
2015	0.979	24.884	0.896	0.190	32.172	0.034
2016	0.991	25.187	0.946	0.129	21.879	0.016
2017	0.989	25.140	0.940	0.143	24.170	0.020
2018	0.975	24.788	0.920	0.128	21.779	0.016

Taula 9: Matriu amb els resultats dels diferents anys

En aquesta taula s'ha pogut observar que, en la dimensió 1, els quatre anys tenen un CTR molt similar, pel voltant del 25%, indicant que tots estan representats amb una contribució similar. Això ens indica que tots els anys tenen una contribució igual a l'hora d'explicar la dimensió. També podem veure un alt valor en el quadrat del cosinus, tenint en compte que el valor màxim és 1. Aquest valor ens indica que la representació dels diferents anys en aquesta dimensió té una alta qualitat.

El cas de la segona dimensió ha estat diferent, Es pot veure que l'any 2015 té un CTR superior a la resta d'anys amb un 32%, és a dir, aquest any té quasi una tercera part de la contribució de la dimensió 2. En els quadrats dels cosinus s'ha vist que aquesta dimensió no gaudeix d'una qualitat gaire bona, mostrant uns valors molt pròxims a 0.

Les figures que es veuen a continuació són una mostra més clara del que s'ha dit anteriorment.

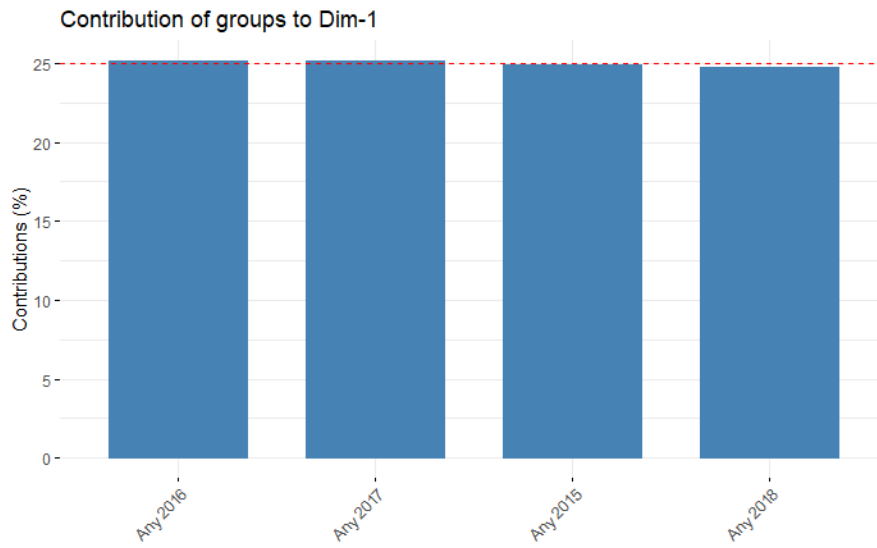


Figura 17: Contribució de cada any en la dimensió 1

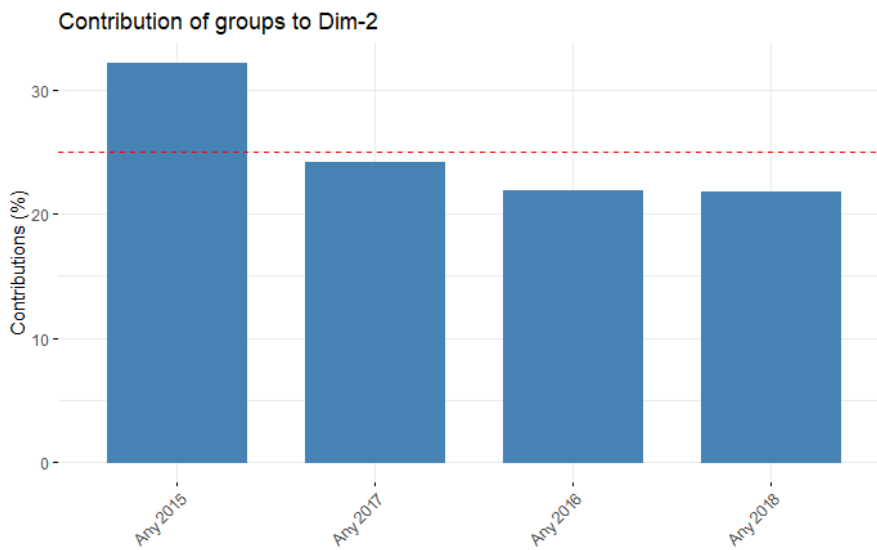


Figura 18: Contribució de cada any en la dimensió 2

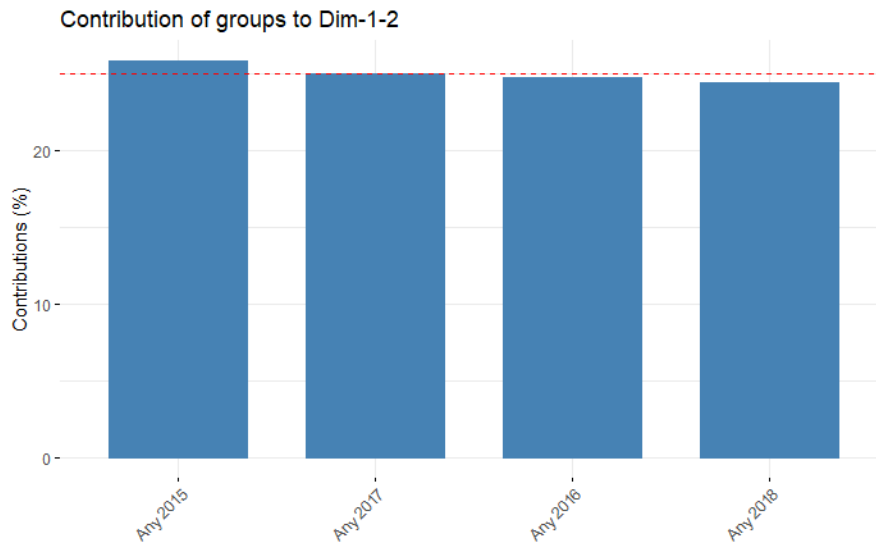


Figura 19: Contribució de cada any en les dimensions 1 i 2

Les figures anteriors ens mostren les contribucions de cada grup en una o més dimensions. Els grups que s'han estudiat han estat els diferents anys dels quals es té informació. En l'eix X es mostra el grup que es representa en la columna de sobre i en l'eix Y la contribució dels diferents grups. En el gràfic es pot veure una línia roja discontinua que ens mostra el valor mitjà de les contribucions dels grups.

En la figura 17 es pot veure la contribució dels diferents anys per a explicar la primera dimensió. El percentatge de contribució de cada any està al voltant del 25%, és a dir, tots els anys tenen gairebé la mateixa contribució per a l'explicació d'aquesta variable.

En la figura 18, s'ha pogut observar la contribució dels diferents anys a l'hora d'explicar la segona dimensió. Es pot veure que, a diferència del cas anterior, en aquest cas sí que hi ha una diferència notable entre els diferents anys, tenint l'any 2015 una contribució considerablement superior a la dels altres anys.

En la figura 19 es mostra la contribució dels anys per explicar les dues primeres dimensions, les dues úniques que utilitzem en el MFA. Es pot veure a simple vista que l'any 2015 té una contribució lleugerament superior a la resta d'anys, tot i que la diferència és mínima.

Tot i mostrar la contribució de les dues dimensions, es pot veure que el gràfic té una forma molt similar a la de la figura 17, que és la que ens explicava la contribució dels diferents anys en la primera dimensió. Això és degut al fet que la primera dimensió té un pes molt superior a la segona.

En aquest apartat s'ha estudiat el comportament de les dades durant els primers anys per a les dues primeres dimensions. Els colors que s'han utilitzat en cada any en el gràfic que es pot veure a continuació s'ha reutilitzat en gràfics que es veuran més avant.

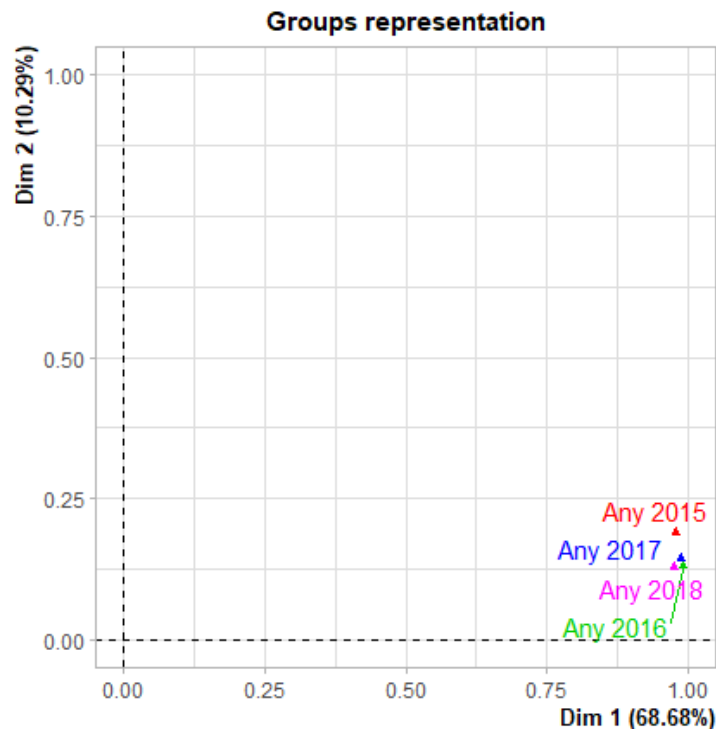


Figura 20: Representació dels grups en dues dimensions

En la figura 20 s'ha il·lustrat la correlació entre els grups i les dimensions. Es pot veure l'eix de les X representat per la dimensió 1 i l'eix de les Y representat per la dimensió 2. En la gràfica s'observa que els grups estan representats per triangles sòlids. Si un grup té un valor de coordenades gran en una dimensió determinada, aquesta dimensió també està present al grup de variables. Si dos grups tenen coordenades grans i properes en diverses dimensions, introdueixen estructures similars als individus.

En aquest gràfic s'observa que, en la primera dimensió, les coordenades dels quatre grups són gairebé iguals. Això ens indica que els quatre anys contribueixen de manera similar a la primera dimensió. La segona dimensió, en canvi, mostra una lleugera diferència entre els diferents anys, mostrant l'any 2016 lleugerament per sobre de la resta.

4.1.3. Contribució de cada individu a les dimensions

Un cop s'ha acabat d'observar la contribució dels grups per a les diferents dimensions, s'ha fet una taula on es mostra la contribució d'alguns dels individus (països) de l'estudi.

	<i>Dim 1</i>	<i>CTR</i>	<i>Cos²</i>	<i>Dim 2</i>	<i>CTR</i>	<i>Cos²</i>
<i>Àustria</i>	-1.954	3.129	0.909	0.016	0.001	0.000
<i>Bèlgica</i>	-2.098	3.609	0.921	0.079	0.034	0.001
<i>Bulgària</i>	2.413	4.772	0.760	-0.261	0.373	0.009
<i>Croàcia</i>	2.787	6.367	0.830	-1.050	6.037	0.118
<i>Xipre</i>	2.231	4.081	0.436	0.663	2.402	0.038
<i>Rep. Txeca</i>	-0.631	0.326	0.108	1.481	12.001	0.594
<i>Dinamarca</i>	-2.402	4.729	0.949	-0.184	0.185	0.006
<i>Estònia</i>	-0.671	0.369	0.278	0.821	3.687	0.416
<i>Finlàndia</i>	-2.385	4.663	0.956	-0.048	0.012	0.000
<i>França</i>	-1.519	1.891	0.578	0.455	1.135	0.052
<i>Alemanya</i>	-1.921	3.025	0.883	0.367	0.737	0.032
<i>Grècia</i>	3.231	8.559	0.831	-0.385	0.811	0.012
<i>Hongria</i>	1.280	1.342	0.685	0.290	0.46	0.035
<i>Islàndia</i>	-1.232	1.244	0.387	-1.466	11.763	0.548
<i>Irlanda</i>	-1.486	1.811	0.344	1.017	5.658	0.161
<i>Itàlia</i>	2.312	4.382	0.776	-0.591	1.911	0.051
<i>Letònia</i>	1.604	2.110	0.511	-0.356	0.693	0.251
<i>Lituània</i>	2.513	5.180	0.736	-0.771	3.256	0.693
<i>Luxemburg</i>	-0.701	0.402	0.246	-0.910	4.529	0.415
<i>Malta</i>	0.110	0.010	0.003	0.724	2.864	0.139
<i>Països Baixos</i>	-2.353	4.540	0.975	-0.169	0.156	0.050
<i>Noruega</i>	-2.450	4.919	0.943	-0.338	0.625	0.018
<i>Polònia</i>	2.251	4.153	0.759	0.577	1.820	0.050
<i>Portugal</i>	0.400	0.131	0.061	0.542	1.605	0.112
<i>Romania</i>	4.175	14.291	0.862	1.097	6.580	0.059
<i>Eslovàquia</i>	1.334	1.459	0.242	2.172	25.802	0.642
<i>Eslovènia</i>	-0.469	0.181	0.164	-0.538	1.586	0.216
<i>Espanya</i>	0.317	0.082	0.065	-0.539	1.593	0.189
<i>Suècia</i>	-2.459	4.955	0.946	-0.316	0.546	0.156

Regne Unit	-1.988	3.241	0.812	0.422	0.974	0.366
------------	--------	-------	-------	-------	-------	-------

Taula 10 : Matriu amb els resultats dels diferents països

En aquesta taula es mostra la mateixa informació que en l'anterior, però en compte d'agrupar-se per anys, s'ha mostrat els resultats per als 10 primers països que hi havia en la nostra base (s'ha mostrat 10 països perquè, si es mostraven tots, la taula resultava massa gran). Es pot veure que en les dues dimensions, tant en el CTR com en el quadrat del cosinus, els valors varien molt entre observacions.

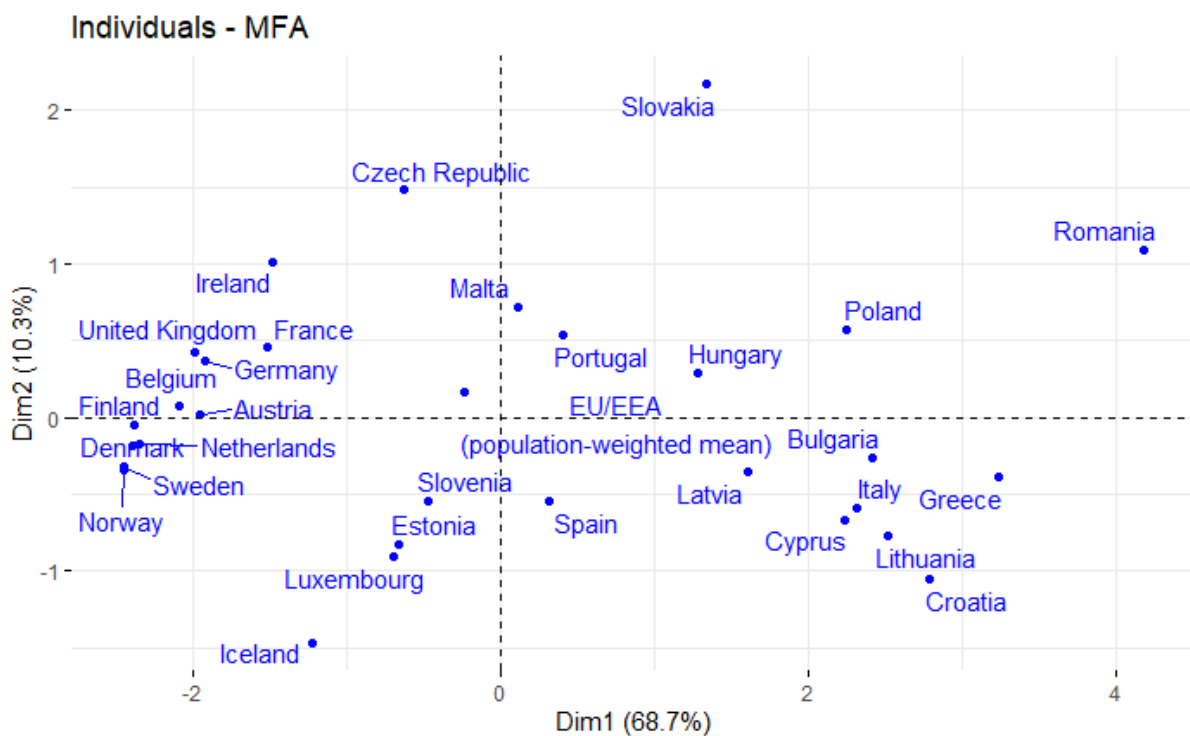


Figura 21: Distribució dels països per a les dues dimensions

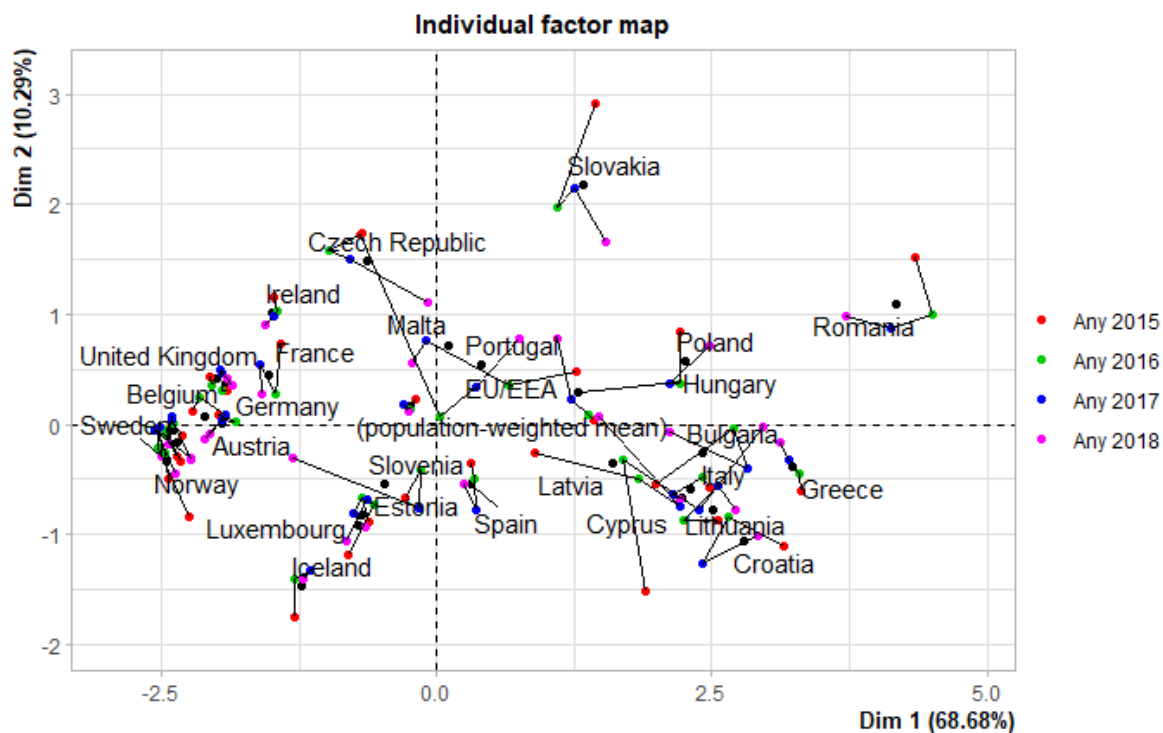


Figura 22: Evolució dels països en els anys per a les dues dimensions

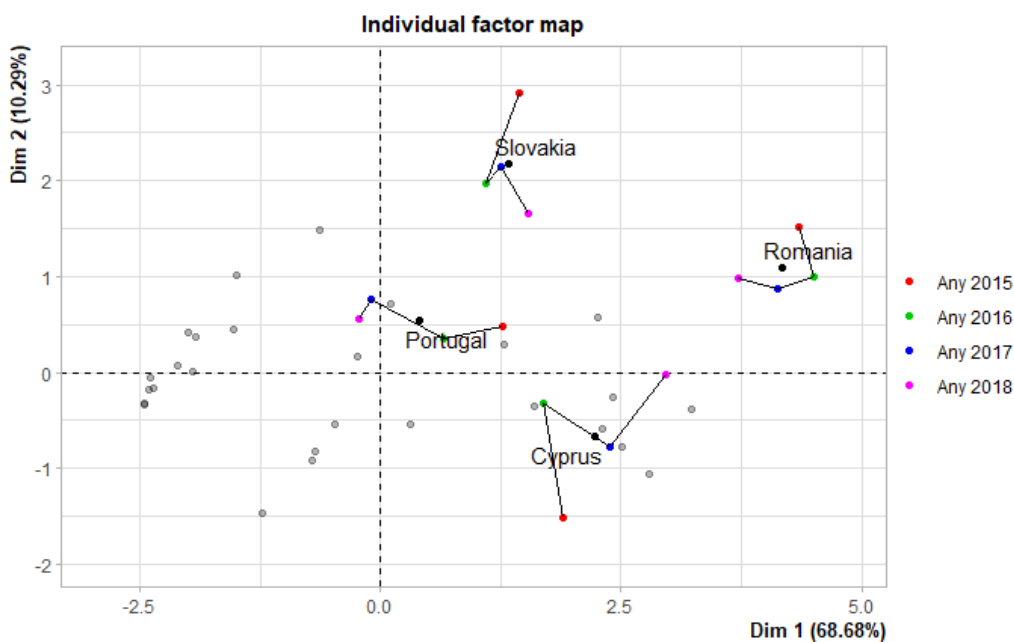


Figura 23: Evolucions particulars en els anys per a les dues dimensions

En la figura 21 es mostra la correlació entre els individus (països) i les dimensions. Es pot veure que, generalment, els països d'Europa Nord tenen una conducta similar en la 1^a dimensió, estant gairebé tots a la part esquerra del gràfic, igual que els països d'Europa occidental.

En el cas dels països situats a l'est d'Europa passa el contrari al de les zones anteriors, tenint una relació en la 1^a dimensió, però, en aquest cas, situant-se a la dreta de la línia de 0. Tampoc sembla que hi hagi una relació per aquests països en la 2^a dimensió.

En el cas d'Europa sud, es pot veure que tots els països estan per sota de la línia del 0 en la 2^a dimensió menys Malta i Portugal. També s'ha observat que estan molt dispersos en la 1^a dimensió.

En la figura 22 s'observen els països estudiats en la mateixa posició que en el gràfic anterior, però, a més a més, també es veu l'evolució d'aquests durant els anys 2015, 2016, 2017 i 2018, representats cadascun d'aquests amb un color diferent. Com hi ha alguns països que se superposen entre si no es pot apreciar correctament l'evolució de molts d'ells, per aquest motiu, s'ha fet un altre gràfic mostrant alguns comportaments particulars.

En la figura 23 es pot observar informació que ja està visible en l'anterior, però, en aquest cas, es pot observar el comportament de Xipre, Portugal, Romania i Eslovàquia. Tenint en compte que el punt roig/taronja és l'any 2015 i el rosa és l'any 2018, es pot observar que, en el cas d'Eslovàquia, hi ha hagut un decreixement en el valor de la segona dimensió mentre que, en la primera, no ha canviat en el pas del temps. En el cas de Portugal s'ha vist un creixement considerable en la 1^a dimensió mentre que en la 2^a no hi ha variat pràcticament. En el Cas de Xipre hi ha hagut un creixement amb el pas del temps, tant en la primera com en la segona dimensió. A Romania, en canvi, ha passat el contrari, és a dir, hi ha hagut un decreixement tant en la primera com en la segona dimensió, tot i que aquests decreixements no han resultat gaire grans.

A continuació es mostra la contribució de les variables quantitatives a les dimensions 1 i 2. Primerament es mostra un conjunt de gràfics, el primer mostrant la contribució de les variables quantitatives en la 1^a dimensió, el segon mostrant la contribució de les variables quantitatives en la 2^a dimensió i, el tercer i últim, mostrant la contribució de les variables quantitatives en les dues dimensions conjuntes. Després es mostraran una taula amb les contribucions de les variables per a les dues primeres dimensions i, després, una altra on es mostren les correlacions entre variables i dimensions.

4.1.4. Contribució de cada variable a les dimensions

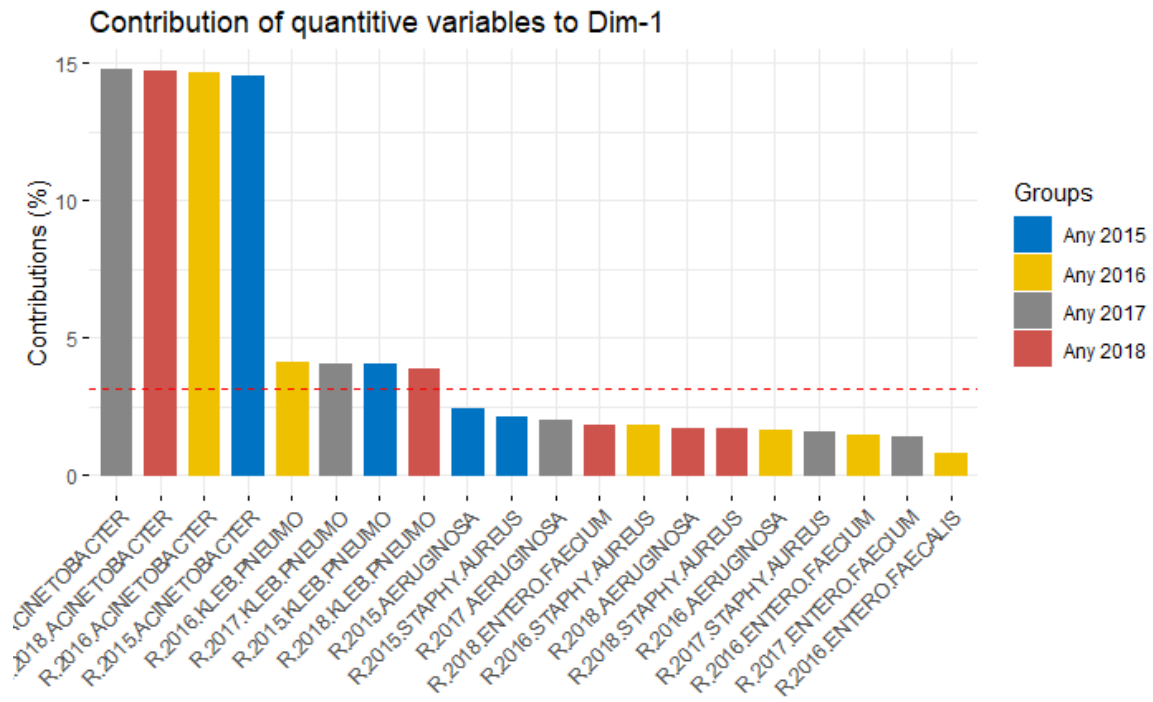


Figura 24: Contribució de les variables quantitatives en la dimensió 1

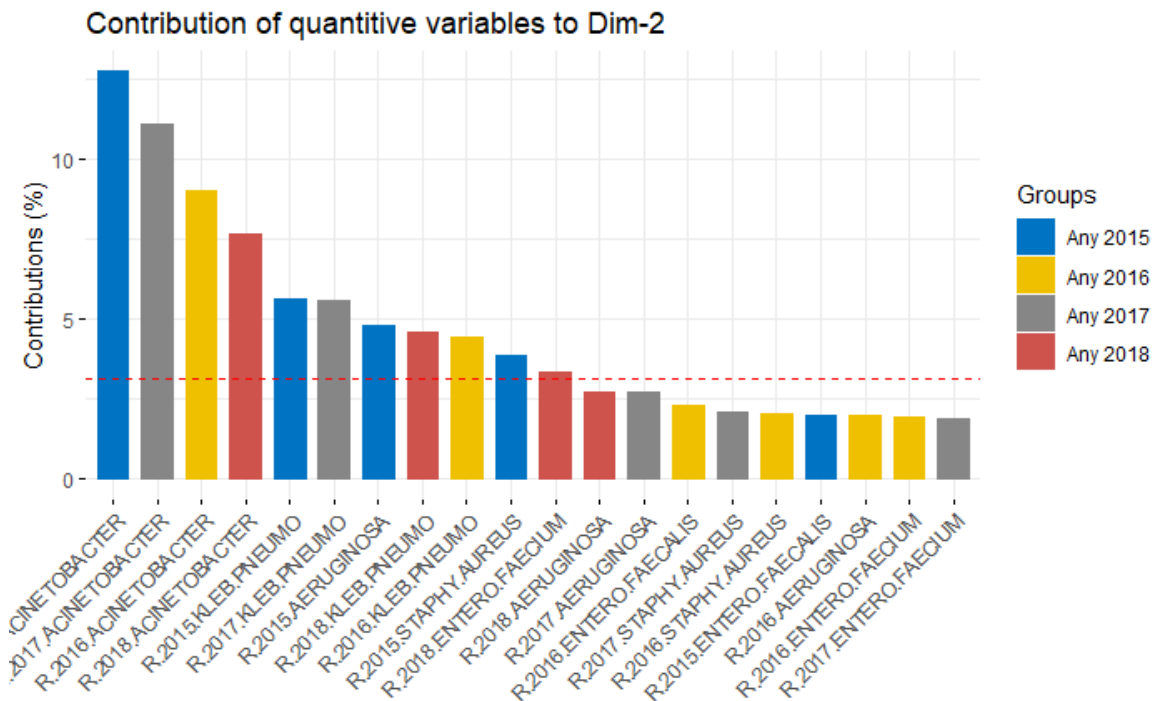


Figura 25: Contribució de les variables quantitatives en la dimensió 2

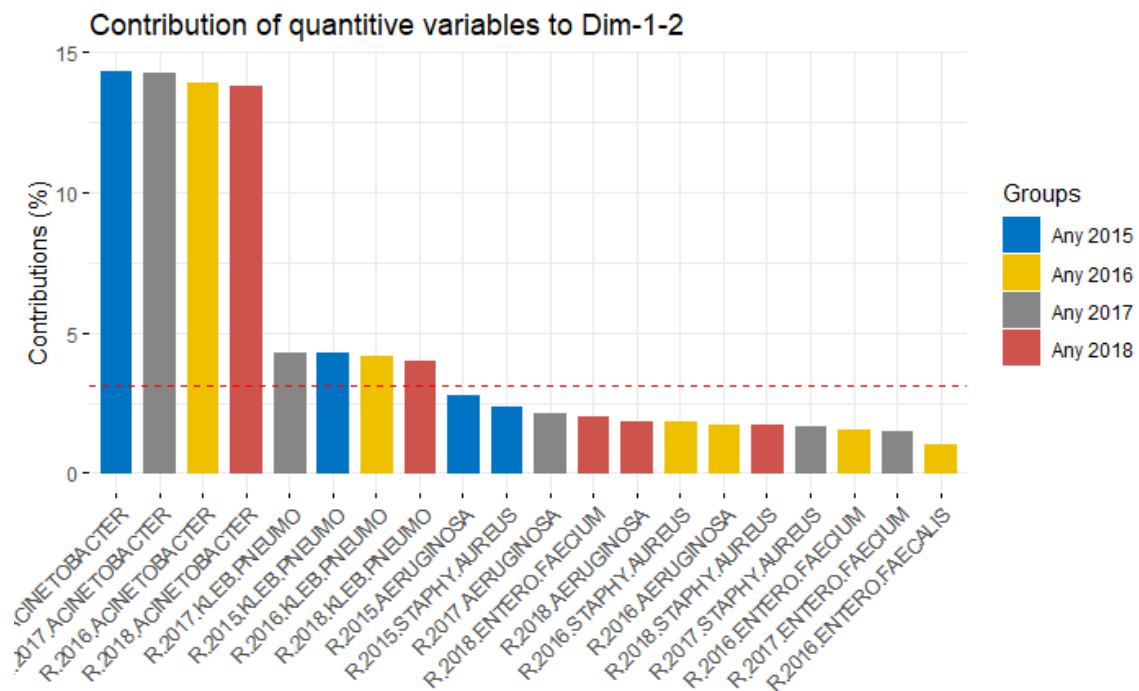


Figura 26: Contribució de les variables quantitatives en les dimensions 1 i 2

	<i>Dimensió 1</i>	<i>Dimensió 2</i>
<i>Acinetobacter 2017</i>	14.745	11.067
<i>Acinetobacter 2018</i>	14.703	7.658
<i>Acinetobacter 2016</i>	14.633	9.037
<i>Acinetobacter 2015</i>	14.524	12.737
<i>Klebsiella pneumoniae 2016</i>	4.132	4.424
<i>Klebsiella pneumoniae 2017</i>	4.075	5.591
<i>Klebsiella pneumoniae 2015</i>	4.053	5.656
<i>Klebsiella pneumoniae 2018</i>	3.901	4.571
<i>Aeruginosa 2015</i>	2.443	4.820
<i>Staphylococcus aureus 2015</i>	2.141	3.850

Taula 11: Contribucions de les variables quantitatives en les dimensions 1 i 2

	<i>Dimensió 1</i>	<i>Dimensió 2</i>
<i>Acinetobacter</i> 2018	0.9487	-0.2648
<i>Acinetobacter</i> 2016	0.9474	-0.2882
<i>Acinetobacter</i> 2017	0.9400	-0.3152
<i>Acinetobacter</i> 2018	0.9254	-0.3355
<i>Klebsiella pneumoniae</i> 2018	0.8481	0.3554
<i>Klebsiella pneumoniae</i> 2016	0.8480	0.3396
<i>Klebsiella pneumoniae</i> 2017	0.8475	0.3843
<i>Klebsiella pneumoniae</i> 2015	0.8214	0.3755
<i>Aeruginosa</i> 2018	0.8173	0.3980
<i>Aeruginosa</i> 2018	0.7964	0.3579

Taula 12: Correlacions entre variables i dimensions

En les figures 21, 22 i 23 s'ha pogut observar els gràfics de contribució de les variables quantitatives, per a la dimensió 1 en el cas de la figura 21, en la dimensió 2 en el cas de la figura 22 i en les dues dimensions ja esmentades en el gràfic 23. En tots tres gràfics s'ha utilitzat el color blau per a les variables del grup pertanyent a l'any 2015, el groc per a les variables del grup pertanyent a l'any 2016, el gris per a les variables pertanyents a l'any 2017 i el roig per al conjunt de variables del grup pertanyent a l'any 2018. S'ha destacat això, ja que, anteriorment, els colors associats a cada grup havia estat diferent. També es pot veure en tots 3 gràfics una línia roja discontinua que ens indica el valor mitjà esperat.

En la figura 21 es pot veure que hi ha quatre variables amb una contribució molt superior a la resta i aquestes quatre variables han estat la resistència del bacteri *Acinetobacter* en els quatre anys d'estudi. Després tenim quatre variables més per sobre de la mitja, on passa el mateix que en el cas de l'*Acinetobacter* però, en aquest cas, amb la *Klebsiella pneumoniae*. Després, totes les variables estan per sota la mitja i deixen de seguir aquest comportament. Es pot observar que en aquesta primera dimensió no s'observa una diferència de contribució a simple vista, ja que, tant en l'*Acinetobacter* com la *Klebsiella pneumoniae*, tenen contribucions molt similars entre els diferents anys i, en la resta del gràfic, tampoc sembla que hi hagi un gran predomini per part de cap any en concret.

En la figura 22 es pot veure que l'*Acinetobacter* torna a ser el bacteri que ocupa les 4 primeres posicions del gràfic, però en aquest cas, la diferència entre anys per aquest bacteri resulta molt superior al cas de la figura 21. Tot seguit es veu que la *Klebsiella pneumoniae* ha estat del bacteri amb una major contribució en aquesta dimensió, però, a diferència de la figura 21, no presenta les 4 següents variables, sinó que entremig també està la resistència de la *Pseudomonas aeruginosa* per l'any 2015. Per acabar amb les variables per sobre de la mitja també hem de nomenar les variables referents a la resistència del *Staphylococcus aureus* en l'any 2015 i de l'*Enterococcus faecium* en l'any 2018. En aquesta figura s'ha vist que, en les variables amb una alta contribució, la contribució de la segona dimensió sembla estar menys condicionada pel tipus de bacteri i més pel grup, tenint una predominança per part de l'any 2015.

En la figura 23, on es mostra la contribució de les variables quantitatives en les dues primeres dimensions, s'ha observat un comportament similar a la figura 21, tenint les variables dels quatre anys per als bacteris *Acinetobacter* i *Klebsiella pneumoniae* per sobre de la mitja. Sembla que en la contribució global de les variables quantitatives, ja que estem treballant amb únicament dues dimensions, no es veu gaire condicionada pel grup.

En la taula 11 es pot veure les contribucions de les 10 variables amb una major contribució en les dimensions 1 i 2. Es pot veure, igual que en la figura de la qual s'acaba de parlar, que les 8 majors variables són les referents als bacteris *Acinetobacter*, amb les quatre variables per sobre del 14% cadascuna, i la *Klebsiella pneumoniae*, amb quatre valors pel voltant del 4. Després van les variables *Aeruginosa* i *Staphylococcus* per l'any 2015. Com les úniques variables que estaven per sobre de la mitja eren les referents a l'*Acinetobacter* i la *Klebsiella pneumoniae*, es pot deduir que el valor de la mitjana de la primera dimensió estava entre 3.901 i 2.443. En la tercera columna de la mateixa taula es pot veure que no estan els valors observats. Això es deu al fet que s'ha introduït les variables amb major contribució per a la primera dimensió i no s'ha tingut en compte la segona per a la tria de les variables que hi hauria a la taula. Tot i això, cal recordar que de les deu variables més contribuents de la dimensió 2, 9 també apareixen en la dimensió 1, però s'haurà de tenir en compte que l'ordre és diferent i, per això, no es mostren de manera ordenada descendent com en la segona columna de la taula 11. Es pot veure que en el cas de les 4 variables de l'*Acinetobacter*, tot i ser les 4 variables més contribuents, tenen una diferència en els valors molt més grans de la que hi havia en la dimensió 1. El mateix passa amb el cas de la *Klebsiella Pneumoniae*, encara que en aquest cas, la variable de la resistència de l'*Aeruginosa* per l'any 2015 estava entre les variables de la *Klebsiella*, és a dir, les 4 variables

de la *Klebsiella* en aquesta 2^a dimensió, a part de tenir una diferència major entre variables, també tenia una variable d'un altre bacteri entremig.

La taula 12 mostra les correlacions de les 10 variables amb una correlació més forta amb les dues dimensions. Es pot veure que les 10 variables amb una correlació més forta tenen una correlació molt gran, estan totes amb una correlació del 80% o superior i, en el cas de les 4 primeres, amb una correlació per sobre del 90%. En el cas de la dimensió 2 les correlacions són molt inferiors a les de la primera, amb una correlació màxima de 0.4447. Aquesta informació no es pot veure en la taula, ja que si introduïem totes les variables ens quedaria una taula amb unes dimensions molt grans. Es pot veure que les correlacions més grans es veuen més condicionades pel bacteri que pel grup referent a l'any.

4.2. Qualitat de representació en les principals dimensions

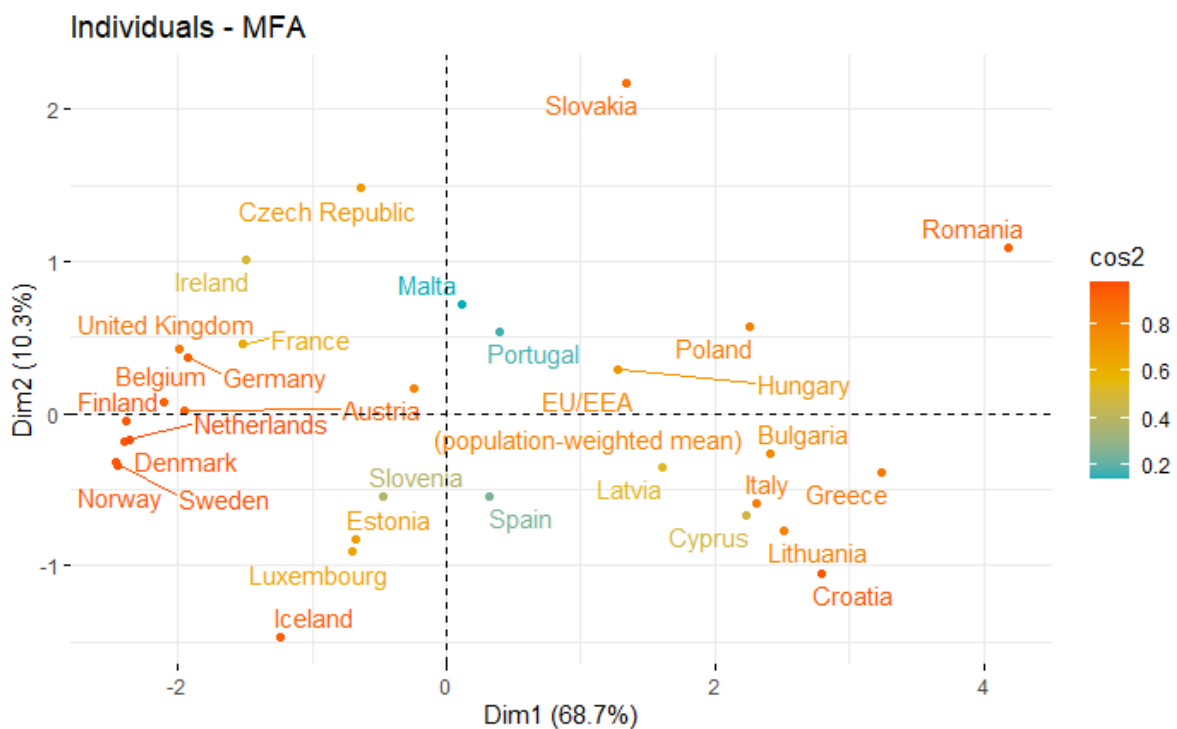


Figura 27: Gràfic dels països amb els \cos^2

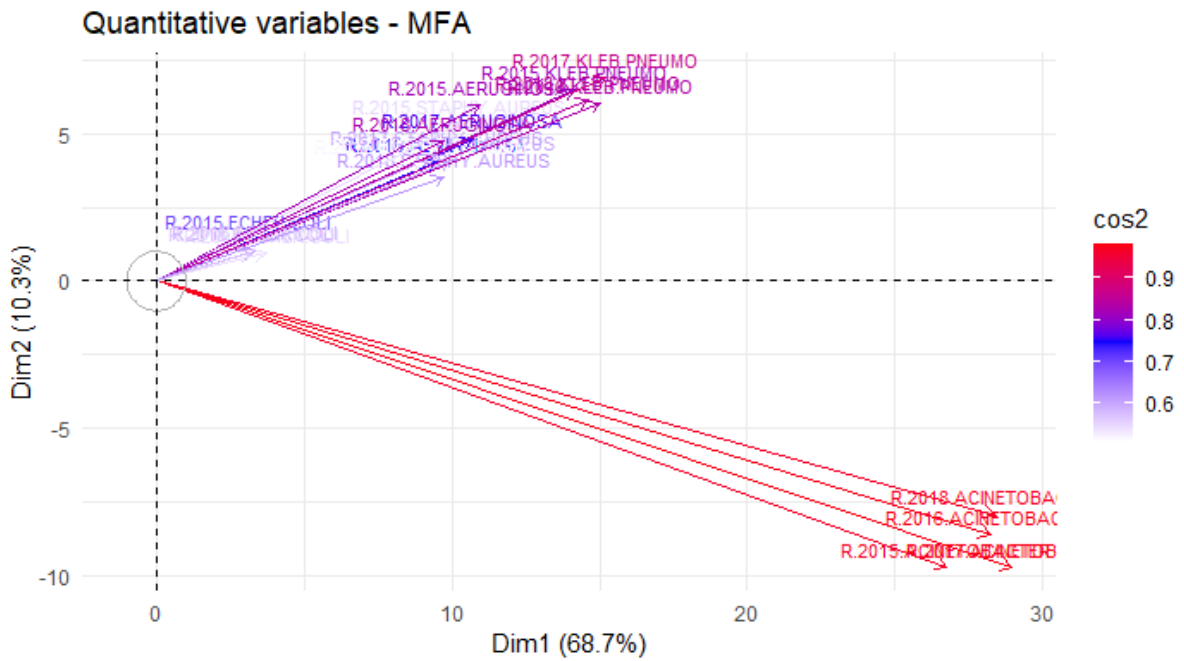


Figura 28: Gràfic de les variables amb $\cos^2 > 0.5$

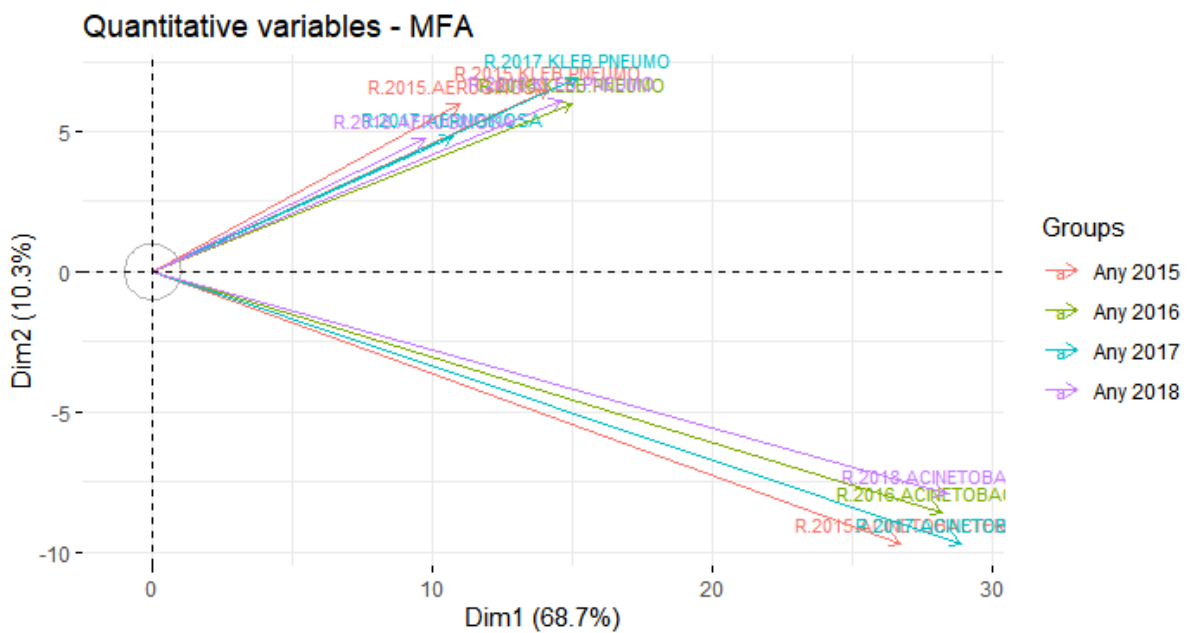


Figura 29: Gràfic de les variables amb $\cos^2 > 0.75$

En la figura 27 es pot observar un gràfic on es veu la distribució dels individus per a les dues dimensions del MFA. A diferència de la figura 18, en aquest gràfic es pot veure la correlació

de cada individu amb les dimensions a través del color, sent el color verd/blau el color que representa una menor correlació i el taronja el que indica una correlació més alta.

Els països amb un valor més baix en el cosinus al quadrat, és a dir, amb una representació amb una qualitat més baixa, han estat Malta i Portugal, mostrant un color que representa un valor pròxim al 0.2. S'ha mirat la resta de països de la seva zona i sembla que la majoria tenen un \cos^2 pròxim a 0.6 pel color. Entre els països d'aquesta zona destaca de manera positiva Croàcia, tenint un color taronja més intens a la resta de països del sud europeu.

Als països de la zona nord del continent es pot veure que els \cos^2 tenen valors variats, tenint des d'individus amb un taronja intens fins a individus amb un groc verdós. No obstant això, el color que més predomina en aquesta zona és el taronja, així que els països de la zona Nord tenen una representació amb una qualitat alta per aquestes dimensions.

En el cas de la zona est, a diferència de les dues zones anteriors, sembla que gairebé tots els països tenen una qualitat similar, ja que els colors dels països oscil·la des d'un taronja no gaire intens fins a un groc ataronjat.

En la zona occidental tots els països gaudeixen d'un alt \cos^2 menys França i Luxemburg, que tenen un color groc, bastant menys intens que la resta de països occidentals.

En la figura 28 s'observa la representació de les diferents variables en les dues dimensions, algunes de les quals han estat rotades. Es pot veure que les variables amb un major valor en el \cos^2 i, per tant, amb una representació de millor qualitat per aquestes dimensions, són les variables del virus *Acinetobacter*. Aquestes variables tenen un \cos^2 molt alt, estan inclús per sobre del 0.9. Això es pot saber observant el color, que és roig per valors alts en el quadrat del cosinus i blau en els valors més baixos. En la part superior del gràfic es mostren totes les variables que tenen un valor superior a 0.5 en el quadrat del cosinus, però tenen un valor inferior a les 4 variables referents al bacteri *Acinetobacter*. Es pot observar que hi ha un gran nombre de variables que superen el valor, això fa que no es pugui apreciar bé la part superior del gràfic.

En el cas de la figura 29, l'exigència ha estat superior, ja que únicament es mostren les variables amb un \cos^2 superior a 0.75. En aquest gràfic el color no depèn del valor del cosinus sinó que el color indica l'any de la variable. Això es deu al fet que en l'anterior figura tenia com a objectiu mostrar-nos quines variables tenien una major qualitat de representació i, en el cas de la figura 29, l'objectiu és veure d'entre totes les variables que tenen una bona representació, quin bloc és el més predominant. Es pot veure que en la part inferior del gràfic hi ha una fletxa a cada color, això es deu al fet que l'*Acinetobacter* té una representació de molt bona qualitat

per a tots els anys. Es pot veure que hi apareixen el mateix nombre de variables per a tots els blocs menys per al bloc de l'any 2016, que n'apareix una menys. Això ens indica que no hi ha un bloc que tingui una qualitat superior als altres en la representació d'aquestes dimensions.

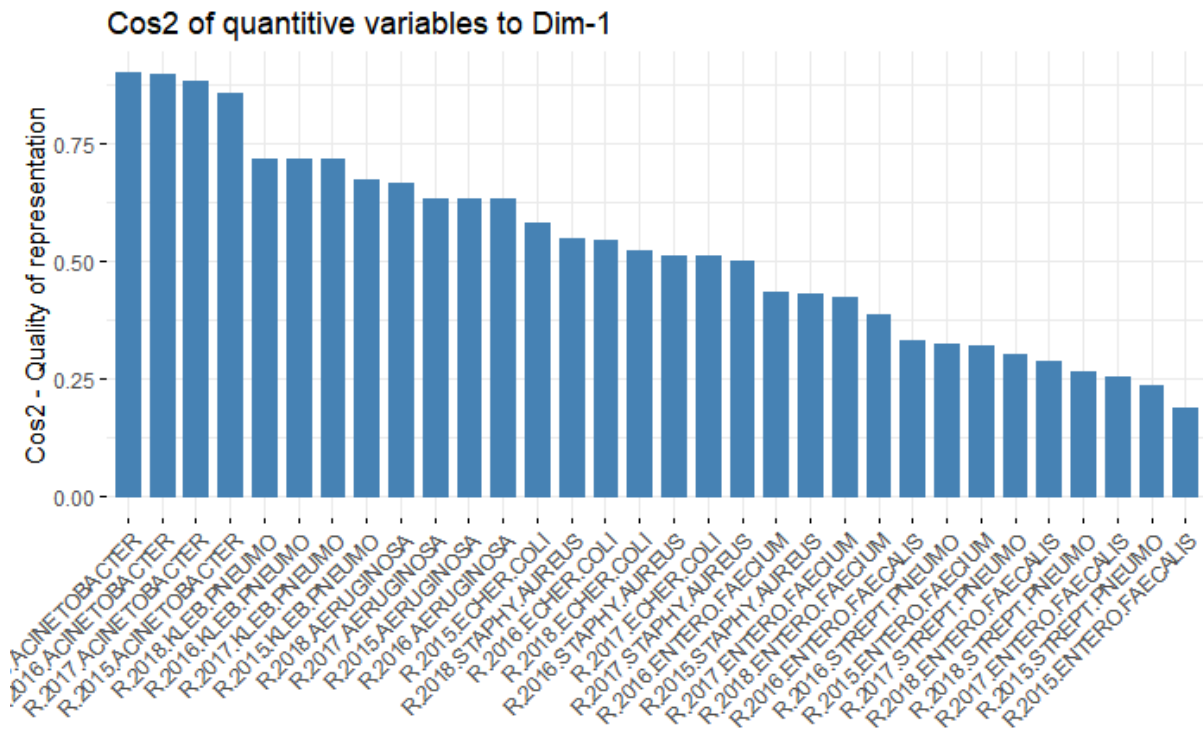


Figura 30: \cos^2 per variables en la dimensió 1

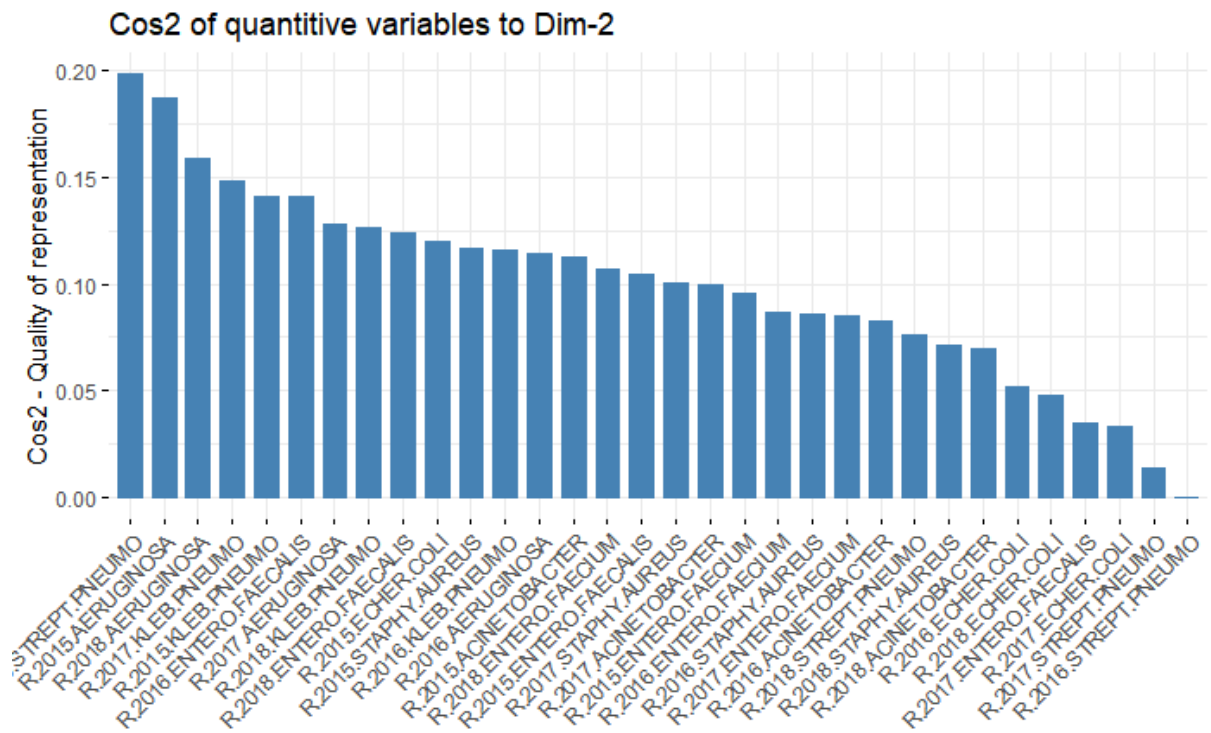


Figura 31: \cos^2 per cada variable en la dimensió 1

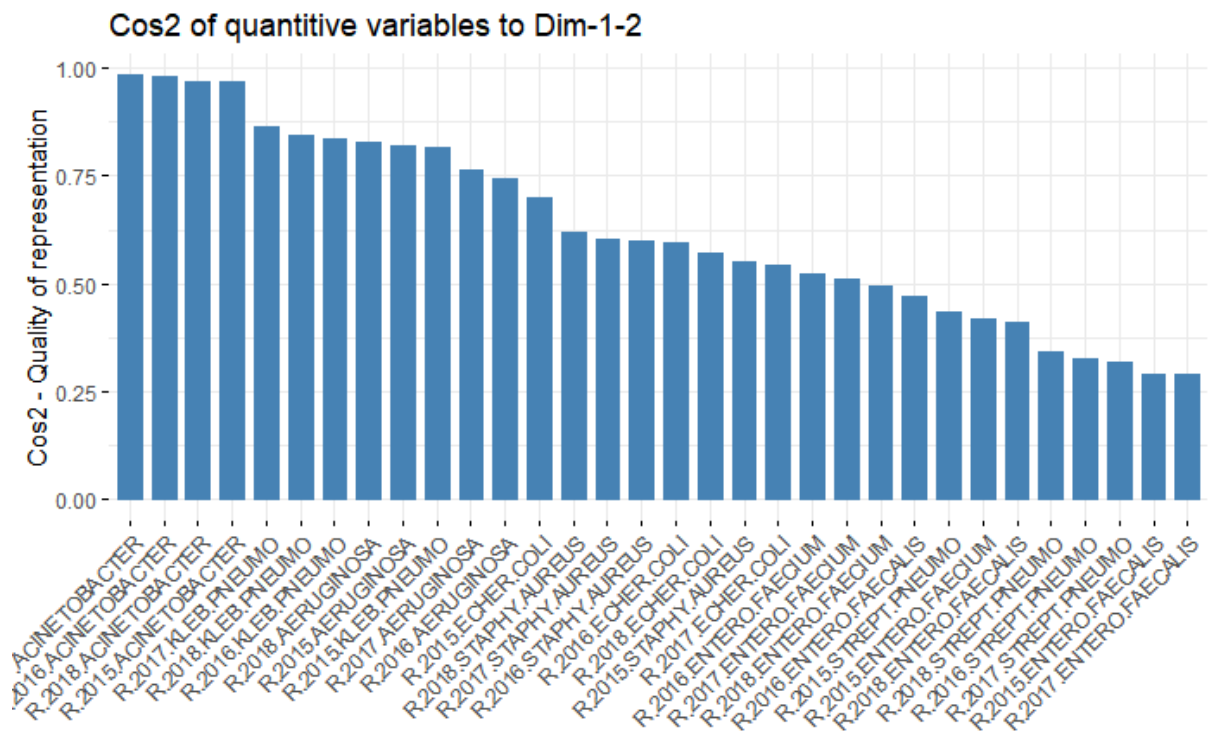


Figura 32: \cos^2 per variables en la dimensions 1 i 2

	<i>DIM. 1</i>	<i>DIM. 2</i>
<i>ACINETOBACTER 2018</i>	0.9000	0.0701
<i>ACINETOBACTER 2016</i>	0.8976	0.0831
<i>ACINETOBACTER 2017</i>	0.8837	0.0993
<i>ACINETOBACTER 2015</i>	0.8564	0.1125
<i>KLEB. PNEUMONIAE 2018</i>	0.7193	0.1263
<i>KLEB. PNEUMONIAE 2016</i>	0.7190	0.1153
<i>KLEB. PNEUMONIAE 2017</i>	0.7181	0.1477
<i>KLEB. PNEUMONIAE 2015</i>	0.6747	0.1410
<i>AERUGINOSA 2018</i>	0.6680	0.1584
<i>AERUGINOSA 2017</i>	0.6342	0.1280

Taula 13: Cos² per variable

	<i>DIM. 1</i>	<i>DIM. 2</i>
<i>ÀUSTRIA</i>	0.9088	0.0000
<i>BÈLGICA</i>	0.9209	0.0013
<i>BULGÀRIA</i>	0.7604	0.0089
<i>CROÀCIA</i>	0.8298	0.1179
<i>XIPRE</i>	0.4359	0.0385
<i>REPÚBLICA TXECA</i>	0.1077	0.5943
<i>DINAMARCA</i>	0.9486	0.0056
<i>ESTÒNIA</i>	0.2775	0.4158
<i>FINLÀNDIA</i>	0.9557	0.0004
<i>FRANÇA</i>	0.5780	0.0520
<i>ALEMANYA</i>	0.8827	0.0323
<i>GRÈCIA</i>	0.8315	0.0118

<i>HONGRIA</i>	0.6854	0.0353
<i>ISLÀNDIA</i>	0.3871	0.5482
<i>IRLANDA</i>	0.3445	0.1613
<i>ITÀLIA</i>	0.7760	0.0507
<i>LETÒNIA</i>	0.5107	0.0251
<i>LITUÀNIA</i>	0.7361	0.0693
<i>LUXEMBURG</i>	0.2459	0.4147
<i>MALTA</i>	0.0032	0.1396
<i>PAÏSOS BAIXOS</i>	0.9752	0.0050
<i>NORUEGA</i>	0.9426	0.0179
<i>POLÒNIA</i>	0.7590	0.0498
<i>PORTUGAL</i>	0.0612	0.1123
<i>ROMANIA</i>	0.8621	0.0595
<i>ESLOVÀQUIA</i>	0.2422	0.6419
<i>ESLOVÈNIA</i>	0.1641	0.2158
<i>ESPANYA</i>	0.0652	0.1889
<i>SUÈCIA</i>	0.9463	0.0156
<i>REGNE UNIT</i>	0.8122	0.0366

Taula 14: cos² per individu

En la figura 30 apareix el valor del \cos^2 per les diferents variables de tots els blocs per la dimensió 1. Es pot veure en l'eix X el nom de les diferents variables de la base original i en l'eix Y la qualitat de la representació. Un cop més, es mostra el bacteri *Acinetobacter* amb una qualitat de representació superior a la resta de bacteris. També es pot veure que no hi ha un any que mostri una representació clarament superior a la resta.

En la figura 31 es pot veure la qualitat de representació de les diferents variables per a la segona dimensió. En aquesta dimensió es pot observar que el bloc de l'any 2015 possiblement és millor

als altres, ja que no hi ha cap variable d'aquest bloc en la part dreta del gràfic, quasi totes les variables estan a la part esquerra i algunes també al mig.

Si es comparen les figures 30 i 31 es pot veure que les variables tenen una qualitat de representació superior en la primera dimensió que en la segona. Això es veu en l'escala, mentre en la primera dimensió les variables tenen una representació des del 25% fins a més del 75%, en la segona dimensió no hi ha cap variable que arribi al 20%.

Si s'observa la figura 32 es pot veure la qualitat de representació per a les dues variables. En aquest gràfic es pot observar, un cop més, que no hi ha una qualitat de representació superior per part de cap bloc.

En la taula 13 s'observa el valor del \cos^2 de les variables amb una major qualitat de representació per a la primera dimensió. En aquesta taula es pot veure el que s'ha dit anteriorment és cert, mentre que en la dimensió 2 no es veu cap variable que arribi al 0.2, en la primera tots els valors estan per sobre de 0.6. Cal recordar que en aquesta taula no apareixen les variables amb major \cos^2 per a la segona dimensió sinó que les variables s'han triat en funció de la primera dimensió.

En la taula 14 es mostra el \cos^2 dels diferents individus per a les dues dimensions. En aquest cas, també sembla que la dimensió una ofereix una qualitat de representació superior, ja que mentre en la dimensió 1 hi ha valors de \cos^2 superiors a 0.95, en la segona dimensió hi ha únicament una observació que supera el 0.6.

4.3. Resultats del MFA

En l'última part d'aquest apartat, com en el seu nom indica, s'interpreten els resultats obtinguts en el MFA. Primerament s'ha vist que la primera dimensió explica un 68.7% de la variància del model i la segona dimensió un 10.3%. Treballant amb aquestes dues dimensions s'ha explicat un 78.97%, un valor més que acceptable per a poder dur a terme el MFA amb dues dimensions. Després s'ha vist que no hi ha cap diferència entre els diferents blocs en la primera dimensió i diferències molt petites entre els mateixos blocs per la segona dimensió. Per al conjunt de les dues dimensions en les quals s'ha treballat tampoc s'han vist diferències. En l'apartat on s'ha vist la contribució de cada variable per a les dimensions s'ha pogut veure que, en el cas de la contribució per a la dimensió 1, la contribució no depenia del bloc sinó del bacteri. En la segona dimensió ha passat el mateix, encara que semblava que estaven d'una

manera lleugerament més aleatòria, ja que no estaven tan agrupats els diferents bacteris entre si com en la primera dimensió. En la contribució de les variables per a les dues dimensions, com era d'esperar, també s'ha observat que l'ordre de la contribució de les variables estava més vinculada al bacteri de la variable que a l'any de la dada.

Finalment, s'ha vist que la qualitat de representació en els individus ha estat relacionada amb la zona dels països. Primer s'hi ha vis que els països de la zona sud d'Europa no mostraven una bona qualitat mentre que les zones nord i occidental mostren una qualitat molt bona en la representació. Els països de la zona est sembla que tenen una qualitat inferior a les dues zones anteriors, però amb una qualitat major que la zona sud.

5. Conclusió

Els objectius principals d'aquest estudi han estat estudiar l'evolució de la Resistència Antimicrobiana (RAM) en els països de la Unió Europea i de l'Espai Econòmic Europeu (EU/EEE) en els darrers anys i analitzar la situació actual d'aquests. S'ha dut a terme una sèrie de tècniques d'anàlisi multivariant amb les dades extretes de l'estudi *Surveillance of antimicrobial resistance in Europe 2018*.

Per tal d'analitzar la situació actual s'ha dut a terme un Anàlisi de Components Principals (PCA) de les dades de l'any 2018, any més recent del qual es tenia informació. Un cop seleccionat un nombre de factors amb el que s'ha vist oportú treballar i s'ha dut a terme l'anàlisi, s'ha descobert que hi ha una relació entre la situació actual i la situació geogràfica del país. Els països de les zones nord i occidental són els que tenen una millor situació actual, tenint una RAM menor als països de les altres zones. En la zona sud generalment s'ha mostrat una RAM molt alta en comparació a les anteriors zones, però en el bacteri *Enterococcus faecium* té una RAM inferior a les altres zones. Els països situats a la zona est d'Europa han mostrat una RAM alta en tots els bacteris, especialment en l'*Enterococcus faecium* i *Acinetobacter*.

Per tal d'estudiar l'evolució de la RAM s'ha dut a terme un Anàlisi factorial Múltiple (MFA). Primerament, s'ha separat en blocs, on cada bloc representava un any i després s'ha dut a terme l'anàlisi. Un cop dut a terme l'anàlisi, no s'ha observat cap diferència entre els diferents blocs, és a dir, no hi ha cap evidència per a poder afirmar que hi ha diferències en la RAM entre els diferents anys.

En conclusió, s'ha observat un patró de comportament en la RAM on els països de les zones nord i occidental han mostrat una RAM inferior als països de les zones sud i est i, per tant, una situació millor. Respecte a l'evolució dels bacteris, no s'ha vist que els valors de la RAM variï amb el pas del temps, és a dir, no sembla que el pas del temps afecti la RAM.

6. Bibliografia i Webgrafia

Addinsoft. *Análisis factorial múltiple (MFA) en Excel*. Centro de ayuda XLSTAT. [consulta el 05.01.2021]. Disponible a: < <https://help.xlstat.com/s/article/analisis-factorial-multiple-mfa-en-excel?language=es> >

Amat Rodrigo, Joaquin. *Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE. plots*. Ciencia de datos.net. Publicat el 18.06.2018. [Consulta: 14.12.2020]. Disponible a: < https://www.cienciadedatos.net/documentos/35_principal_component_analysis >

Bernal, Juan Jesús; Martínez, Soledad; Sánchez, Juan Francisco. Modelización de los factores más importantes que caracterizan un sitio en la red. Universidad Politécnica de Cartagena - XII Jornadas de ASEPUMA. [consulta el 05.12.2020]. Disponible a: < https://www.um.es/asepuma04/comunica/bernal_martinez_sanchez.pdf >

Bolaños, Luis. *Análisis factorial*. RStudio. Publicat el 11.03.2020. [consulta el 29.12.2020]. Disponible a: < https://rstudio-pubs-static.s3.amazonaws.com/584396_130454ceaaf946d786f35df57bdf659.html >

Cassini A, Hogberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. *Attributable deaths and disability-adjusted lifeyears caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis*. Publicat per: Lancet Infectious Diseases. 2019 Jan;19(1):56-66. [Consulta el 20.10.2020]. Disponible a: < [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(18\)30605-4/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(18)30605-4/fulltext)>

Centre Europeu per a la Prevenció i Control de Malalties (ECDC). *Surveillance of antimicrobial resistance in Europe 2018*, Estocolm, November 2019. [Consulta: varies]. Disponible a: < <https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2018.pdf> >

Chávez, Marcelo. *Introducción a los Métodos Multivariantes. Análisis Factorial y de Componentes Principales*. RPubs. Publicat al febrer de 2017. [Consulta el 15.01.2021]. Disponible a: < https://rpubs.com/marcelo-chavez/multivariado_1 >

Comissió Europea. *A European One Health Action Plan against Antimicrobial Resistance (AMR)*. Brussel·les. 2017. [Consulta el 20.10.2020] Disponible a:< https://ec.europa.eu/health/amr/sites/amr/files/amr_action_plan_2017_en.pdf >

Factominer. *Multiple Factor Analysis*. [Consulta el 02.01.2021]. Disponible a: < <http://factominer.free.fr/factomethods/multiple-factor-analysis.html> >

Gil Martínez, Cristina. *Análisis de Componentes principales*. RPubs. Publicat al juny de 2018. [Consulta el 15.12.2020]. Disponible a: < https://rpubs.com/Cristina_Gil/PCA >

Hayden, Luke. *Principal Component Analysis in R*. Publicat el 09.08.2018. [Consulta el 21.12.2020]. Disponible a: < <https://www.datacamp.com/community/tutorials/pca-analysis-r> >

Husson, F.; Lê, S.; Pagès, J. *Exploratory Multivariate Analysis by Example using R*. Editat per CRC Press en 2017. ISBN: 978-1-1381-9634-6.[consulta el 05.01.2021]. Disponible a: < <https://books.google.es> >

Kassambara. *MFA - Multiple Factor Analysis in R: Essentials*. Statistical tools for high-throughput data analysis (STHDA). [Consulta el 05.01.2021]. Disponible a: < <http://sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/116-mfa-multiple-factor-analysis-in-r-essentials/> >

Linh Ngo. *How to read PCA biplots and scree plots* [en línia]. Publicat el 18.06.2018. [Consulta el 19.11.2020]. Disponible a: < <https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/> >

López-Roldán, Pedro; Fachelli, Sandra. *Metodología de la Investigación Social Cuantitativa*. Editat per UAB en febrer de 2015. [Consulta el 20.11.2020]. Disponible a: < https://ddd.uab.cat/pub/caplli/2015/142928/metinvsocua_cap3-11a2016v3.pdf >

Madigan, Michael T.; Martinko, John M.; Bender, Kelly S.; Buckley, Daniel H., Stahl, David A. *Brock, biología de los microorganismos*. Editat per Pearson Education SA en 2015. ISBN: 978-84-9035-279-3 [Consulta el 15.11.2020].

Maguina, Misael; Gómez, Doris. *Análisis factorial*. RPubS. Publicat el 11.07.2018. [Consulta el 28.12.2020]. Disponible a: < <https://rpubs.com/misael892/Factorial> >

Marin, Juan Miguel. *Análisis Multivariante*. Universidad Carlos III. [Consulta el 20.12.2020]. Disponible a: < <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/AMult.html> >

Merck Sharp & Dohme de España, S.A. *¿Qué es una infección bacteriana?* [Consulta: varies]. Disponible a: < <https://www.msdsalud.es/cuidar-en/infecciones/infecciones-bacterianas/es-una-infeccion-bacteriana.html> >

Merck Sharp & Dohme de España, S.A. *Manual MSD para el público en general*. [Consulta: varies]. Disponible a: < <https://www.msmanuals.com/es-es/hogar> >

Minitab. *Interpretar todos los estadísticos y gráficas para Análisis de componentes principales*. [Consulta el 26.11.2020]. Disponible a: < <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/> >

National Human Genome Research Institute. *El glosario hablado de términos genéticos*. [Consulta el 30.11.2020]. Disponible a: < <https://www.genome.gov/es/genetics-glossary> >

Organització per a la Cooperació i el Desenvolupament Econòmic (OCDE) i Centre Europeu per a la Prevenció i Control de Malalties (ECDC). *Antimicrobial Resistance. Tackling the burden in the European Union. Briefing note for EU/EEA countries*. Paris: OECD 2019. [Consulta el 19.11.2020]. Disponible a: < <https://www.oecd.org/health/health-systems/AMR-Tackling-the-Burden-in-the-EU-OECD-ECDC-Briefing-Note-2019.pdf> >

Organització Mundial de la Salut: *Global Action Plan on Antimicrobial Resistance*. Ginebra 2015. [Consulta el 20.11.2020]. Disponible a: <https://apps.who.int/iris/bitstream/handle/10665/193736/9789241509763_eng.pdf?sequence=1>

Prats, Miguel. *Microbiología y Parasitología Médicas*. Editat per Editorial Médica Panamericana en 2012. ISBN: 978-84-9835-429-4. 18a edició [Consulta el 05.12.20]. Disponible a: <<http://eap-books.club/readonline/?item=8498354293&lan=en>>

R Package Documentation. *PcaNA: Classical or robust Principal Components for incomplete data*. [Consulta el 12.12.2020]. Disponible a: <<https://rdrr.io/cran/rrcovNA/man/PcaNA.html>>

Terrádez, Manuel. *Análisis de componentes principales*. Universitat Oberta de Catalunya. [Consulta el 25.12.2020]. Disponible a: <http://web.archive.org/web/20091229050509/http://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf>

Vu, Vincent Vu. *ggbiplot v0.55*. Rdocumentation. [Consulta el 29.11.2020]. Disponible a: <www.rdocumentation.org/packages/ggbiplot/versions/0.55>

7. Glossari

ACP Anàlisi de Components Principals. Tècnica utilitzada per reduir la dimensionalitat d'un conjunt de dades.

Carboxismes Microcompartiments bacterians que contenen enzims implicats en la fixació del carboni.

Deteriorament sistèmic Deteriorament del sistema.

EARS-Net European Antimicrobial Resistance Surveillance Network. Base de dades central i completa per a la Unió Europea que se centra en set patògens bacterians diferents.

ECDC European Centre for Disease Prevention and Control. Agència de la Unió Europea que s'ocupa de lluitar contra les malalties infeccioses.

Flagels Orgànul amb forma de fuet que usen molts organismes unicel·lulars i uns pocs de pluricel·lulars.

EEE Espai Econòmic Europeu. Conjunt de països que participen en el mercat únic europeu.

Endospores Estructura inactiva, resistent, no reproductiva produïda per una petita quantitat de bacteris de la divisió Firmicute.

IQR Rang Intercuartílic. Mesura de dispersió estadística igual a la diferència entre tercer i el primer quartil.

Magnetosomes Estructures membranoses presents en bacteris magnetotàctics.

MCA Multiple Correspondence Analysis. Tècnica d'anàlisi de dades categòriques que s'utilitza per detectar i representar estructures subjacents en un conjunt de dades.

MFA Multiple Factor Analysis. Tècnica de reducció de dades utilitzada per tal d'explicar les correlacions entre les variables observades.

PCA Principal Component Analysis. Tècnica utilitzada per reduir la dimensionalitat d'un conjunt de dades.

Pili Estructura en forma de pèl més curta i fina que els flagels que es troben a la superfície de molts bacteris.

Plasmidi Molècula d'ADN circular de doble cadena pròpia dels procariotes, que pot existir i replicar-se independentment del cromosoma.

pH Potencial de l'hidrògen. Mesura d'acidesa o alcalinitat d'una dissolució.

Prosteca Prolongaciones del citoplasma y de la pared celular que se pueden encontrar en algunos tipos de bacteria.

Quimioheterotrof Individu que utilitza compostos químics en els seus processos metabòlics com a font d'energia.

RAM Resistència antimicrobiana. Resistència als antibiòtics per part d'un bacteri.

UE Unió Europea. Unió econòmica i política de 27 Estats independents, localitzats majoritàriament a Europa.

Annex codi

```
#install.packages("rrcovNA")
library("rrcovNA")
#install.packages("FactoMineR")
library("FactoMineR")
#install.packages("factoextra")
library("factoextra")

#dades<-read.table("Ressist_Comb_anys_ord.csv",header = T, sep = ";", dec="," )
#dades<-read.table("Ressist_Comb_anys.csv", header = T, sep = ";", dec="," )
dades <- read.csv2("C:/Users/guill/OneDrive/Desktop/Ressist_Comb_anys.csv")
a2015<-seq(2,33,4)
a2016<-seq(3,33,4)
a2017<-seq(4,33,4)
a2018<-seq(5,33,4)
dades<-dades[,c(1,a2015,a2016,a2017,a2018)]

row.names(dades)<-dades[,1]
dades<-dades[,-1]
View(dades)

###Descriptiva

summary(dades)
summary(dades)
colnames(dades)
attach(mtcars)

##Descriptiva 2015
summary(dades[,c(1:8)])
##Descriptiva 2016
summary(dades[,c(9:16)])
##Descriptiva 2017
summary(dades[,c(17:24)])
##Descriptiva 2018
summary(dades[,c(25:32)])

#Escherichia Coli
boxplot(dades$R.2015.ECHER.COLI, dades$R.2016.ECHER.COLI,
dades$R.2017.ECHER.COLI, dades$R.2018.ECHER.COLI, main = "Escherichia Coli",
names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab = "Resistance")

which(dades$R.2015.ECHER.COLI>16)
dades$COUNTRY[27]
which(dades$R.2016.ECHER.COLI>16)
dades$COUNTRY[3]
which(dades$R.2017.ECHER.COLI>16)
dades$COUNTRY[c(3,27)]
which(dades$R.2018.ECHER.COLI>16)
```

```
dades$COUNTRY[c(3,27)]
```

```
hist(dades$R.2015.ECHER.COLI, breaks = seq(from = min(dades$R.2015.ECHER.COLI),  
to= max(dades$R.2015.ECHER.COLI), length.out = 11), main = "Escherichia Coli 2015",  
xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2016.ECHER.COLI, breaks = seq(from = min(dades$R.2016.ECHER.COLI),  
to= max(dades$R.2016.ECHER.COLI), length.out = 11), main = "Escherichia Coli 2016",  
xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.ECHER.COLI, breaks = seq(from = min(dades$R.2017.ECHER.COLI),  
to= max(dades$R.2017.ECHER.COLI), length.out = 11), main = "Escherichia Coli 2017",  
xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2018.ECHER.COLI, breaks = seq(from = min(dades$R.2018.ECHER.COLI),  
to= max(dades$R.2018.ECHER.COLI), length.out = 11), main = "Escherichia Coli 2018",  
xlab = "Resistencia", ylab="Freq")
```

```
#Klebsiella pneumoniae
```

```
boxplot(dades$R.2015.KLEB.PNEUMO, dades$R.2016.KLEB.PNEUMO,  
dades$R.2017.KLEB.PNEUMO, dades$R.2018.KLEB.PNEUMO, main = "Klebsiella  
pneumoniae", names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab = "Resistance")
```

```
hist(dades$R.2015.KLEB.PNEUMO, breaks = seq(from =  
min(dades$R.2015.KLEB.PNEUMO), to= max(dades$R.2015.KLEB.PNEUMO), length.out  
= 11), main = "Klebsiella pneumoniae 2015", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2016.KLEB.PNEUMO, breaks = seq(from =  
min(dades$R.2016.KLEB.PNEUMO), to= max(dades$R.2016.KLEB.PNEUMO), length.out  
= 11), main = "Klebsiella pneumoniae 2016", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.KLEB.PNEUMO, breaks = seq(from =  
min(dades$R.2017.KLEB.PNEUMO), to= max(dades$R.2017.KLEB.PNEUMO), length.out  
= 11), main = "Klebsiella pneumoniae 2017", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2018.KLEB.PNEUMO, breaks = seq(from =  
min(dades$R.2018.KLEB.PNEUMO), to= max(dades$R.2018.KLEB.PNEUMO), length.out  
= 11), main = "Klebsiella pneumoniae 2018", xlab = "Resistencia", ylab="Freq")
```

```
#Pseudomonas aeruginosa
```

```
boxplot(dades$R.2015.AERUGINOSA, dades$R.2016.AERUGINOSA,  
dades$R.2017.AERUGINOSA, dades$R.2018.AERUGINOSA, main = "Pseudomonas  
aeruginosa", names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab = "Resistance")
```

```
which(dades$R.2015.AERUGINOSA>42)
```

```
dades$COUNTRY[26]
```

```
which(dades$R.2016.AERUGINOSA>40)
```

```
dades$COUNTRY[26]
```

```
which(dades$R.2017.AERUGINOSA>35)
```

```
dades$COUNTRY[c(18,26,27)]
```

```
which(dades$R.2018.AERUGINOSA>40)
dades$COUNTRY[26]
```

```
hist(dades$R.2015.AERUGINOSA, breaks = seq(from =
min(dades$R.2015.AERUGINOSA), to= max(dades$R.2015.AERUGINOSA), length.out =
11), main = "Pseudomonas aeruginosa 2015", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2016.AERUGINOSA, breaks = seq(from =
min(dades$R.2016.AERUGINOSA), to= max(dades$R.2016.AERUGINOSA), length.out =
11), main = "Pseudomonas aeruginosa 2016", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.AERUGINOSA, breaks = seq(from =
min(dades$R.2017.AERUGINOSA), to= max(dades$R.2017.AERUGINOSA), length.out =
11), main = "Pseudomonas aeruginosa 2017", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2018.AERUGINOSA, breaks = seq(from =
min(dades$R.2018.AERUGINOSA), to= max(dades$R.2018.AERUGINOSA), length.out =
11), main = "Pseudomonas aeruginosa 2018", xlab = "Resistencia", ylab="Freq")
```

```
#Acinetobacter species
```

```
boxplot(dades$R.2015.ACINETOBACTER,
dades$R.2016.ACINETOBACTER,dades$R.2017.ACINETOBACTER,dades$R.2018.ACIN
ETOBACTER, main = "Acinetobacter species", names = c("2015", "2016", "2017", "2018"),
xlab = "any", ylab = "Resistance")
```

```
hist(dades$R.2015.ACINETOBACTER, breaks = seq(from =
dades$R.2015.ACINETOBACTER[which.min(dades$R.2015.ACINETOBACTER)],
to=dades$R.2015.ACINETOBACTER[which.max(dades$R.2015.ACINETOBACTER)],
length.out = 11), main = "Acinetobacter species 2015", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2016.ACINETOBACTER, breaks = seq(from =
dades$R.2016.ACINETOBACTER[which.min(dades$R.2016.ACINETOBACTER)],
to=dades$R.2016.ACINETOBACTER[which.max(dades$R.2016.ACINETOBACTER)],
length.out = 11), main = " Acinetobacter species 2016", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.ACINETOBACTER, breaks = seq(from =
dades$R.2017.ACINETOBACTER[which.min(dades$R.2017.ACINETOBACTER)],
to=dades$R.2017.ACINETOBACTER[which.max(dades$R.2017.ACINETOBACTER)],
length.out = 11), main = "Acinetobacter species 2017", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2018.ACINETOBACTER, breaks = seq(from =
dades$R.2018.ACINETOBACTER[which.min(dades$R.2018.ACINETOBACTER)],
to=dades$R.2018.ACINETOBACTER[which.max(dades$R.2018.ACINETOBACTER)],
length.out = 11), main = "Acinetobacter species 2018", xlab = "Resistencia", ylab="Freq")
```

```
#Streptococcus pneumoniae
```

```
boxplot(dades$R.2015.STREPT.PNEUMO, dades$R.2016.STREPT.PNEUMO,
dades$R.2017.STREPT.PNEUMO, dades$R.2018.STREPT.PNEUMO, main =
"Streptococcus pneumoniae", names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab
= "Resistance")
```

```

which(dades$R.2015.STREPT.PNEUMO>25)
which(dades$R.2016.STREPT.PNEUMO>25)
dades$COUNTRY[c(5, 26)]
which(dades$R.2017.STREPT.PNEUMO>30)
dades$COUNTRY[5]
which(dades$R.2018.STREPT.PNEUMO>25)
dades$COUNTRY[26]

```

```

hist(dades$R.2015.STREPT.PNEUMO, breaks = seq(from =
dades$R.2015.STREPT.PNEUMO[which.min(dades$R.2015.STREPT.PNEUMO)],
to=dades$R.2015.STREPT.PNEUMO[which.max(dades$R.2015.STREPT.PNEUMO)],
length.out = 11), main = "Streptococcus pneumoniae 2015", xlab = "Resistencia",
ylab="Freq")

```

```

hist(dades$R.2016.STREPT.PNEUMO, breaks = seq(from =
dades$R.2016.STREPT.PNEUMO[which.min(dades$R.2016.STREPT.PNEUMO)],
to=dades$R.2016.STREPT.PNEUMO[which.max(dades$R.2016.STREPT.PNEUMO)],
length.out = 11), main = "Streptococcus pneumoniae 2016", xlab = "Resistencia",
ylab="Freq")

```

```

hist(dades$R.2017.STREPT.PNEUMO, breaks = seq(from =
dades$R.2017.STREPT.PNEUMO[which.min(dades$R.2017.STREPT.PNEUMO)],
to=dades$R.2017.STREPT.PNEUMO[which.max(dades$R.2017.STREPT.PNEUMO)],
length.out = 11), main = "Streptococcus pneumoniae 2017", xlab = "Resistencia",
ylab="Freq")

```

```

hist(dades$R.2018.STREPT.PNEUMO, breaks = seq(from =
dades$R.2018.STREPT.PNEUMO[which.min(dades$R.2018.STREPT.PNEUMO)],
to=dades$R.2018.STREPT.PNEUMO[which.max(dades$R.2018.STREPT.PNEUMO)],
length.out = 11), main = "Streptococcus pneumoniae 2018", xlab = "Resistencia",
ylab="Freq")

```

```

#Staphylococcus aureus

```

```

boxplot(dades$R.2015.STAPHY.AUREUS, dades$R.2016.STAPHY.AUREUS,
dades$R.2017.STAPHY.AUREUS, dades$R.2018.STAPHY.AUREUS, main =
"Staphylococcus aureus", names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab =
"Resistance")

```

```

which(dades$R.2015.STAPHY.AUREUS>50)
dades$COUNTRY[26]
which(dades$R.2016.STAPHY.AUREUS>51)
which(dades$R.2017.STAPHY.AUREUS>45)
which(dades$R.2018.STAPHY.AUREUS>45)

```

```

hist(dades$R.2015.STAPHY.AUREUS, breaks = seq(from =
dades$R.2015.STAPHY.AUREUS[which.min(dades$R.2015.STAPHY.AUREUS)],
to=dades$R.2015.STAPHY.AUREUS[which.max(dades$R.2015.STAPHY.AUREUS)],
length.out = 11), main = "Staphylococcus aureus 2015", xlab = "Resistencia", ylab="Freq")

```

```

hist(dades$R.2016.STAPHY.AUREUS, breaks = seq(from =
dades$R.2016.STAPHY.AUREUS[which.min(dades$R.2016.STAPHY.AUREUS)],

```

```
to=dades$R.2016.STAPHY.AUREUS[which.max(dades$R.2016.STAPHY.AUREUS)],
length.out = 11), main = "Staphylococcus aureus 2016", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.STAPHY.AUREUS, breaks = seq(from =
dades$R.2017.STAPHY.AUREUS[which.min(dades$R.2017.STAPHY.AUREUS)],
to=dades$R.2017.STAPHY.AUREUS[which.max(dades$R.2017.STAPHY.AUREUS)],
length.out = 11), main = "Staphylococcus aureus 2017", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.STAPHY.AUREUS, breaks = seq(from =
dades$R.2017.STAPHY.AUREUS[which.min(dades$R.2017.STAPHY.AUREUS)],
to=dades$R.2017.STAPHY.AUREUS[which.max(dades$R.2017.STAPHY.AUREUS)],
length.out = 11), main = "Staphylococcus aureus 2018", xlab = "Resistencia", ylab="Freq")
```

```
#Enterococcus faecalis
```

```
boxplot(dades$R.2015.ENTERO.FAECALIS, dades$R.2016.ENTERO.FAECALIS,
dades$R.2017.ENTERO.FAECALIS, dades$R.2018.ENTERO.FAECALIS, main =
"Enterococcus faecalis", names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab =
"Resistance")
```

```
hist(dades$R.2015.ENTERO.FAECALIS, breaks = seq(from =
dades$R.2015.ENTERO.FAECALIS[which.min(dades$R.2015.ENTERO.FAECALIS)],
to=dades$R.2015.ENTERO.FAECALIS[which.max(dades$R.2015.ENTERO.FAECALIS)],
length.out = 11), main = "Enterococcus faecalis 2015", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2016.ENTERO.FAECALIS, breaks = seq(from =
dades$R.2016.ENTERO.FAECALIS[which.min(dades$R.2016.ENTERO.FAECALIS)],
to=dades$R.2016.ENTERO.FAECALIS[which.max(dades$R.2016.ENTERO.FAECALIS)],
length.out = 11), main = "Enterococcus faecalis 2016", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.ENTERO.FAECALIS, breaks = seq(from =
dades$R.2017.ENTERO.FAECALIS[which.min(dades$R.2017.ENTERO.FAECALIS)],
to=dades$R.2017.ENTERO.FAECALIS[which.max(dades$R.2017.ENTERO.FAECALIS)],
length.out = 11), main = "Enterococcus faecalis 2017", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2018.ENTERO.FAECALIS, breaks = seq(from =
dades$R.2018.ENTERO.FAECALIS[which.min(dades$R.2018.ENTERO.FAECALIS)],
to=dades$R.2018.ENTERO.FAECALIS[which.max(dades$R.2018.ENTERO.FAECALIS)],
length.out = 11), main = "Enterococcus faecalis 2018", xlab = "Resistencia", ylab="Freq")
```

```
#Enterococcus faecium
```

```
boxplot(dades$R.2015.ENTERO.FAECIUM, dades$R.2016.ENTERO.FAECIUM,
dades$R.2017.ENTERO.FAECIUM, dades$R.2018.ENTERO.FAECIUM, main =
"Enterococcus faecium", names = c("2015", "2016", "2017", "2018"), xlab = "any", ylab =
"Resistance")
```

```
which(dades$R.2015.ENTERO.FAECIUM>30)
dades$COUNTRY[26]
```

```
which(dades$R.2016.ENTERO.FAECIUM>48)
which(dades$R.2017.ENTERO.FAECIUM>45)
which(dades$R.2018.ENTERO.FAECIUM>60)
```



```
hist(dades$R.2015.ENTERO.FAECIUM, breaks = seq(from =
dades$R.2015.ENTERO.FAECIUM[which.min(dades$R.2015.ENTERO.FAECIUM)],
to=dades$R.2015.ENTERO.FAECIUM[which.max(dades$R.2015.ENTERO.FAECIUM)],
length.out = 11), main = "Enterococcus faecium 2015", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2016.ENTERO.FAECIUM, breaks = seq(from =
dades$R.2016.ENTERO.FAECIUM[which.min(dades$R.2016.ENTERO.FAECIUM)],
to=dades$R.2016.ENTERO.FAECIUM[which.max(dades$R.2016.ENTERO.FAECIUM)],
length.out = 11), main = "Enterococcus faecium 2016", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2017.ENTERO.FAECIUM, breaks = seq(from =
dades$R.2017.ENTERO.FAECIUM[which.min(dades$R.2017.ENTERO.FAECIUM)],
to=dades$R.2017.ENTERO.FAECIUM[which.max(dades$R.2017.ENTERO.FAECIUM)],
length.out = 11), main = "Enterococcus faecium 2017", xlab = "Resistencia", ylab="Freq")
```

```
hist(dades$R.2018.ENTERO.FAECIUM, breaks = seq(from =
dades$R.2018.ENTERO.FAECIUM[which.min(dades$R.2018.ENTERO.FAECIUM)],
to=dades$R.2018.ENTERO.FAECIUM[which.max(dades$R.2018.ENTERO.FAECIUM)],
length.out = 11), main = "Enterococcus faecium 2018", xlab = "Resistencia", ylab="Freq")
```

```
###PCA
```

```
cor(dades[,c(25:32)], use="pairwise") #com hi ha dades missins utilitzem la opció pairwise
```

```
install.packages("rrcovNA")
```

```
library(rrcovNA)
```

```
dades_pca<-PcaNA(dades[,c(25:32)],scale=T) # per defecte scale=F,center=T
```

```
summary(dades_pca) # Importància de les components (en los primers 4 PCA expliquem casi
el 99% de la variança)
```

```
plot(dades_pca) #grafic dades del PCA
```

```
print(dades_pca, print.x=TRUE) #info pca
```

```
PcaNA(dades[,c(25:32)], method="class") #sdev dels components
```

```
plot(PcaNA(dades[,c(25:32)], k=2)) #grafic
```

```
screplot(dades_pca) #gràfic de variancies
```

```
loadings(dades_pca) #vectors propis
```

```
#install.packages("ggplot2")
```

```
library(ggplot2)
```

```
prop_var <- dades_pca$eigenvalues / sum(dades_pca$eigenvalues)
```

```
prop_var #sd dels components
```

```
prop_var_acum <-cumsum(prop_var)
```

```
prop_var_acum
```

```
ggplot(data = data.frame(prop_var_acum, pc =1:8),
```

```
  aes(x = pc, y = prop_var_acum, group = 1))+
```

```
  geom_point() +
```

```
  geom_line() +
```

```

    geom_label(aes(label = round(prop_var_acum,4))) +
    theme_bw() +
    labs(x = "Component principal",
         y = "Prop. var. explicada acumulada")

##PC1vsPC2

#dades_pca$scores

plot(dades_pca$scores[,1],dades_pca$scores[,2],main='PCA Resistència
2018',xlab='PC1',ylab='PC2',type='n')
text(dades_pca$scores[,1],dades_pca$scores[,2], label=row.names(dades),cex=0.9)
abline(h=0,col='red');abline(v=0,col='red')

biplot(dades_pca,cex=c(0.7,0.9), col =c("blue4" , "black"), choices=c(1,2))
#scale 0 per a que les fletxes estiguin a la mateixa Dim que los components

##pc1vs3

#dades_pca$scores

plot(dades_pca$scores[,1],dades_pca$scores[,3],main='PCA Resistència
2018',xlab='PC1',ylab='PC3',type='n')
text(dades_pca$scores[,1],dades_pca$scores[,3], label=row.names(dades),cex=0.9)
abline(h=0,col='red');abline(v=0,col='red')

biplot(dades_pca,cex=c(0.7,0.9), col =c("blue4" , "black"), choices=c(1,3)) #añadir xlim i
ylim

##pc2vs3

#dades_pca$scores

plot(dades_pca$scores[,2],dades_pca$scores[,3],main='PCA Resistència
2018',xlab='PC2',ylab='PC3',type='n')
text(dades_pca$scores[,2],dades_pca$scores[,3], label=row.names(dades),cex=0.9)
abline(h=0,col='red');abline(v=0,col='red')

biplot(dades_pca,cex=c(0.7,0.9), col =c("blue4" , "black"), choices=c(2,3)) #añadir xlim i
ylim

##Calculs PCA

corre<-cor(dades[,c(25:32)], predict(dades_pca)[,1:3], use="pairwise") #Matriu de
components
corre
varcomp1<-apply(corre^2, 1, sum)
varcomp2<-apply(corre^2, 2, sum) #Var explicada pels components

```

```

round(varcomp1[1],6) == round(corre[1,1]^2 + corre[1,2]^2 + corre[1,3]^2, 6)
round(varcomp1[2],6) == round(corre[2,1]^2 + corre[2,2]^2 + corre[2,3]^2, 6)
round(varcomp1[3],6) == round(corre[3,1]^2 + corre[3,2]^2 + corre[3,3]^2, 6)
round(varcomp1[4],6) == round(corre[4,1]^2 + corre[4,2]^2 + corre[4,3]^2, 6)
round(varcomp1[5],6) == round(corre[5,1]^2 + corre[5,2]^2 + corre[5,3]^2, 6)
round(varcomp1[6],6) == round(corre[6,1]^2 + corre[6,2]^2 + corre[6,3]^2, 6)
round(varcomp1[7],6) == round(corre[7,1]^2 + corre[7,2]^2 + corre[7,3]^2, 6)
round(varcomp1[8],6) == round(corre[8,1]^2 + corre[8,2]^2 + corre[8,3]^2, 6)

round(varcomp2[1],6) ==
round(corre[1,1]^2+corre[2,1]^2+corre[3,1]^2+corre[4,1]^2+corre[5,1]^2+corre[6,1]^2+corre
[7,1]^2+corre[8,1]^2, 6)
round(varcomp2[2],6) ==
round(corre[1,2]^2+corre[2,2]^2+corre[3,2]^2+corre[4,2]^2+corre[5,2]^2+corre[6,2]^2+corre
[7,2]^2+corre[8,2]^2, 6)
round(varcomp2[3],6) ==
round(corre[1,3]^2+corre[2,3]^2+corre[3,3]^2+corre[4,3]^2+corre[5,3]^2+corre[6,3]^2+corre
[7,3]^2+corre[8,3]^2, 6)

rota <- varimax(corre, normalize = T)
rota$loadings

#grafic de components en espai rotat
plot(rota$loadings[,1],rota$loadings[,2],type="n",xlim=c(-1,1),ylim=c(-1,1))
text(rota$loadings[,1],rota$loadings[,2],labels=row.names(rota$loadings))
title(main="Gràfico de components en espai rotat")
abline(h=0, v=0, col = "gray60")
#Podem veure la correlació de les diferents variables per als 2 components principals

plot(rota$loadings[,1],rota$loadings[,3],type="n",xlim=c(-1,1),ylim=c(-1,1))
text(rota$loadings[,1],rota$loadings[,3],labels=row.names(rota$loadings))
title(main="Gràfico de components en espai rotat")
abline(h=0, v=0, col = "gray60")
#Podem veure la correlació de les diferents variables per al 3r component principal

#install.packages("MASS")
library(MASS)
eqscplot(predict(dades_pca)[,1:2],type="n",xlab="PC1" ,ylab="PC2")#Gràfic que usa la
mateixa escala en els dos eixos
text(predict(dades_pca) [,1:2],labels=row.names(dades))
title(main="Estars en 1ª y 2ª components")

###MFA
res.mfa <- MFA(dades,
  group = c(8, 8, 8, 8),
  type = c("c", "c", "c", "c"),
  name.group = c("Any 2015","Any 2016","Any 2017","Any 2018"),
  num.group.sup = NULL,
  graph = FALSE)
summary(res.mfa, ncp=2)

```

```

# Diagrama amb el percentatge de variabilitat explicat per cada dimensiÃ³ del MFA
fviz_screplot(res.mfa,addlabels=T)

# La primera dimensiÃ³ ella sola explica prop d'un 70% de la variabilitat
#i amb la segona prop d'un 80%

#Grups

# ContribuciÃ³ de cada grup a les dimensions

fviz_contrib(res.mfa,'group',axes=1)
fviz_contrib(res.mfa,'group',axes=2)
fviz_contrib(res.mfa,'group',axes=1:2)

#RepresentaciÃ³ dels grups
plot(res.mfa,choix='group')
# Com es pot comprovar sobre la primera dimensiÃ³ els 4 anys estÃ n prÃ cticament
# solapats. Ã% la segona dimensiÃ³ la que separa els anys (i no massa)

#Individus

res.mfa$ind$coord[,c(1, 2)] #Matriu DIM
res.mfa$ind$contrib[,c(1, 2)] #Matriu CTR
res.mfa$ind$cos2[,c(1, 2)] #Matriu Cos2

# Grafic dels diferents paÃ- sos
#
fviz_mfa_ind(res.mfa,palette = 'lancet',repel = TRUE)

plot(res.mfa, choix = "ind", partial="all",habillage="group", chrono=TRUE)

# Alguns comportaments particulars

plot(res.mfa, choix = "ind", partial="all",habillage="group", chrono=TRUE,
      select=c('Romania','Portugal','Slovakia','Cyprus'))

#Variables

# ContribuciÃ³ de les variables quantitatives a les dimensions

fviz_contrib(res.mfa, choice = "quanti.var", axes = 1, top = 20,
             palette = "jco",labels=1)
fviz_contrib(res.mfa, choice = "quanti.var", axes = 2, top = 20,
             palette = "jco",labels=1)
fviz_contrib(res.mfa, choice = "quanti.var", axes = c(1,2), top = 20,
             palette = "jco",labels=1)

```

```

# Taules amb les contribucions per variables
DT::datatable(res.mfa$global.pca$var$contrib,fillContainer = F)

# Taules amb les correlacions per variables
DT::datatable(res.mfa$global.pca$var$cor,fillContainer = F)

#Cos

#cos2
fviz_cos2(res.mfa, choice = "quanti.var", axes = 1)
fviz_cos2(res.mfa, choice = "quanti.var", axes = 2)
fviz_cos2(res.mfa, choice = "quanti.var", axes = 1:2)

# GrÀfic amb el cos2
fviz_mfa_ind(res.mfa, col.ind = "cos2",gradient.cols = c("#00AFBB", "#E7B800",
"#FC4E07"),
            rebel = TRUE)

# Gràfic amb variables amb cos2 > 0.5
fviz_mfa_var(res.mfa, col.var="cos2",labelsize=3,select.var=list(cos2=0.5),shape.var='group')
+
  scale_color_gradient2(low="white", mid="blue",
                        high="red", midpoint=0.75) + theme_minimal()

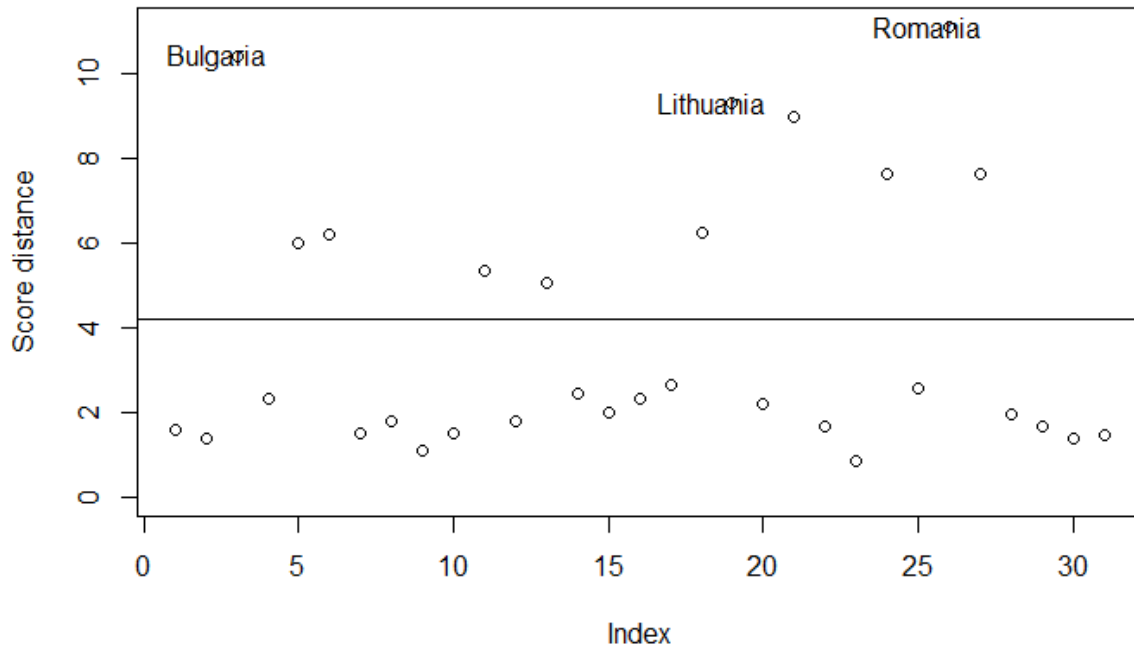
# Gràfic amb variables amb cos2 > 0.75
fviz_mfa_var(res.mfa,select.var=list(cos2=0.75),labelsize=3)

DT::datatable(res.mfa$global.pca$var$cos2,fillContainer = F) # Taula amb el cos^2 per
variable
DT::datatable(res.mfa$global.pca$ind$cos2,fillContainer = F) # Taula amb el cos^2 per
individu

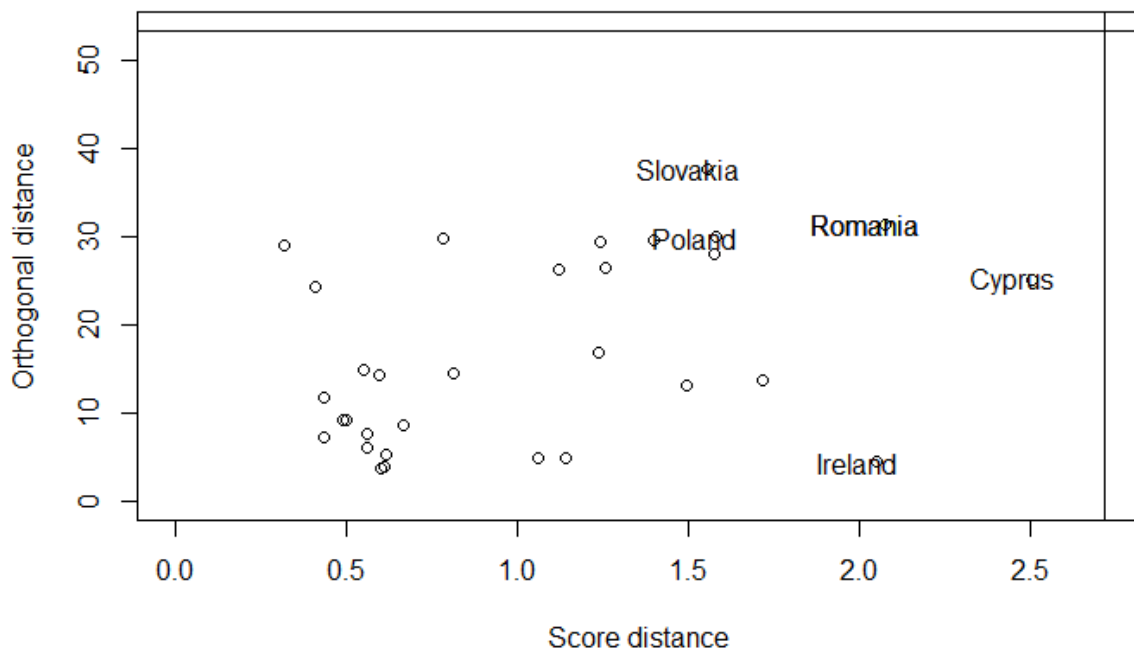
```

Annex d'imatges i taules

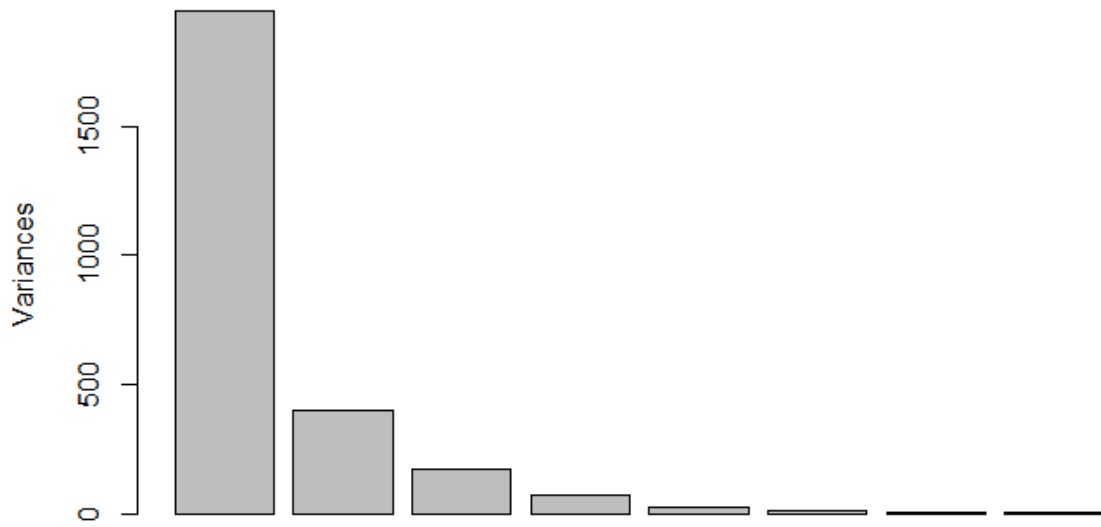
Robust PCA



Robust PCA



getPrcomp(x)



Enterococcus faecium 2018

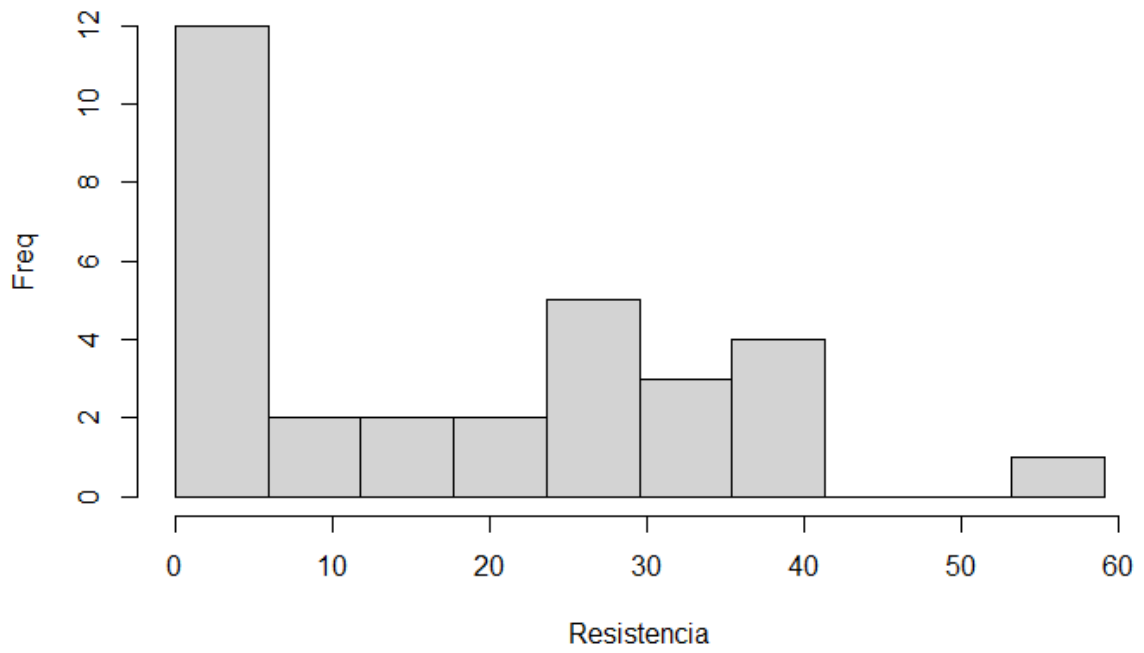
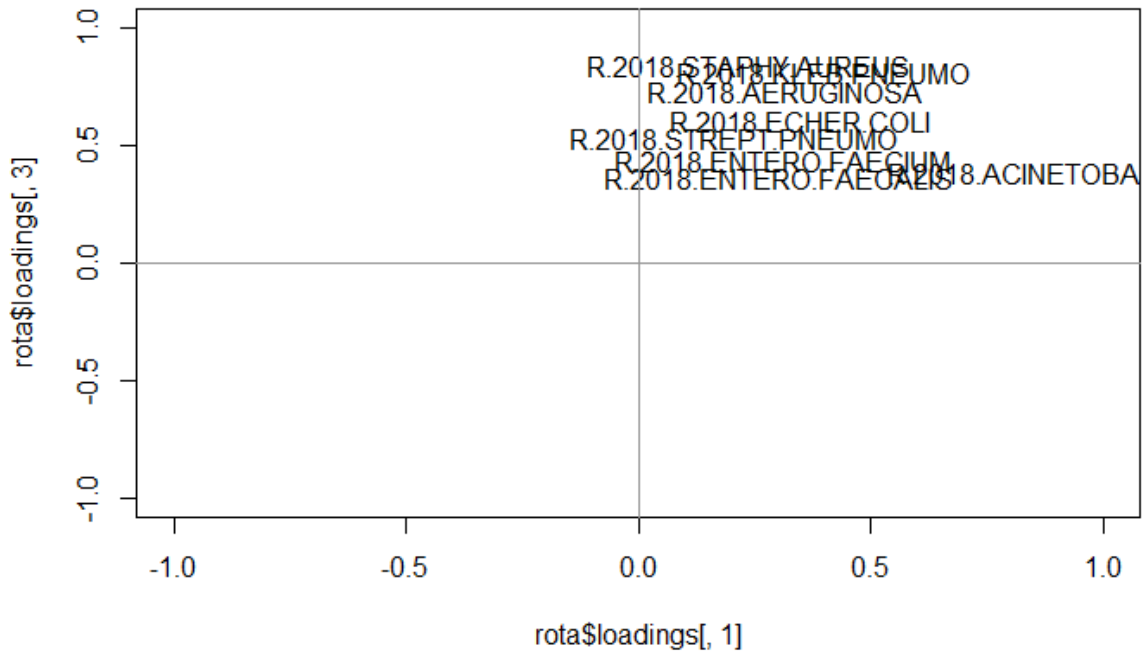


Gráfico de components en espai rotat



Estados en 1ª y 2ª componentes

