

SELECCIÓN DE CARACTERÍSTICAS Y ANÁLISIS DE CORRESPONDENCIAS PARA LA EXTRACCIÓN DE CONOCIMIENTO DE PACIENTES CRÓNICOS

Autora: Natalia Alonso Arteaga
Tutora: Inmaculada Mora Jiménez
Cotutora: Cristina Soguero Ruíz

17 de abril de 2020

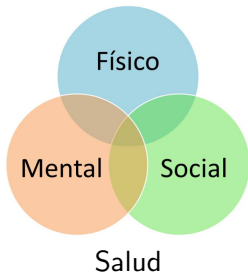
Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 Análisis exploratorio de datos
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple
- 6 Conclusiones y líneas futuras

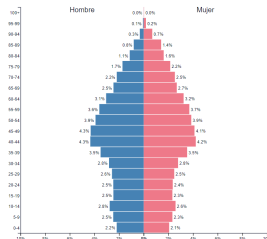
Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 Análisis exploratorio de datos
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple
- 6 Conclusiones y líneas futuras

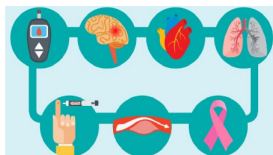
Contexto y motivación



Sistemas sanitarios



Población envejecida



Enfermedades crónicas

Enfermedades crónicas y comorbilidades

Enfermedades de larga
duración y por lo general de
progresión lenta



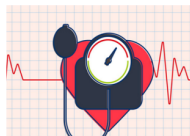
Depresión

Trastorno en el estado anímico, prolongado en el tiempo, y que afecta a la capacidad de enfrentarse a las actividades cotidianas.



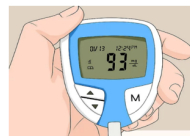
Enfermedad Pulmonar Obstructiva Crónica (EPOC)

Inflamación de los pulmones que obstruye el flujo de aire y que produce dificultad para respirar.



Hipertensión Arterial (HTA)

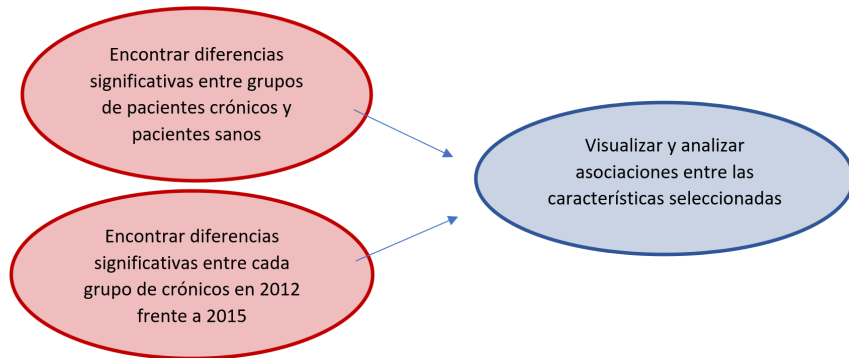
Aumento de la presión arterial o de la fuerza que la sangre ejerce contra las paredes de las arterias.



Diabetes (DM)

Imposibilidad de regular los niveles de azúcar en sangre, pudiendo producir graves consecuencias para el organismo.

Objetivos



Test proporciones + remuestreo *bootstrap*

Análisis de Correspondencias Múltiple

Codificación de diagnósticos y fármacos

- **Clasificación Internacional de Enfermedades (CIE-9):** para la clasificación y codificación de **diagnósticos**, **enfermedades** y lesiones, **procedimientos clínicos**, **factores** que influyen en el estado de salud, causas externas o neoplasias.

250 . 00	V 01 71	E 880 0	M 809 0 / 1
Diabetes Mellitus Tipo II	Varicela	Caída en escalera mecánica	Tumor basocelular

- **Clasificación Anatómica, Terapéutica, Química (ATC):** para codificación de sustancias farmacológicas y **medicamentos**, en base al sistema orgánico sobre el que actúa, el efecto farmacológico, las indicaciones terapéuticas y la estructura química.

Nivel	Significado	Código	Descripción
Nivel 1	Grupo anatómico principal	A	Tracto alimentario y metabolismo
Nivel 2	Subgrupo terapéutico	A10	Fármacos usados en diabetes
Nivel 3	Subgrupo terapéutico farmacológico	A10B	Fármacos hipoglucemiantes orales
Nivel 4	Subgrupo químico-terapéutico	A10BA	Biguanidas
Nivel 5	Principio activo	A10BA02	Metformina



Sistemas de Clasificación de Pacientes: SCP

- Son sistemas que clasifican pacientes en un número determinado de **grupos homogéneos**. Los individuos en un mismo grupo tienen **características clínicas** y un **consumo de recursos** sanitarios similar.
- Un caso particular es el *Clinical Risk Group* (CRG) de la compañía 3M, un SCP poblacional que clasifica **pacientes** en base a sus **características clínicas, demográficas y farmacológicas** a través de los datos recogidos en los **contactos** de los pacientes con el sistema sanitario en un periodo temporal, en nuestro caso de un año.

Estado de salud	Descripción
1	Estado de salud sano
2	Historia de enfermedad aguda significativa
3	Enfermedad crónica menor única
4	Enfermedad crónica menor en diferentes sistemas orgánicos
5	Enfermedad dominante o moderada crónica única
6	Enfermedad significativa crónica en dos o más sistemas orgánicos
7	Enfermedad dominante crónica en múltiples sistemas orgánicos
8	Neoplasias dominantes, metastásicas y complicadas
9	Necesidades sanitarias elevadas

CRG- base	Descripción
3755	Depresión
5133	EPOC
5192	HTA
5424	DM
6313	HTA+DM
7140	HTA+DM+otra cronicidad

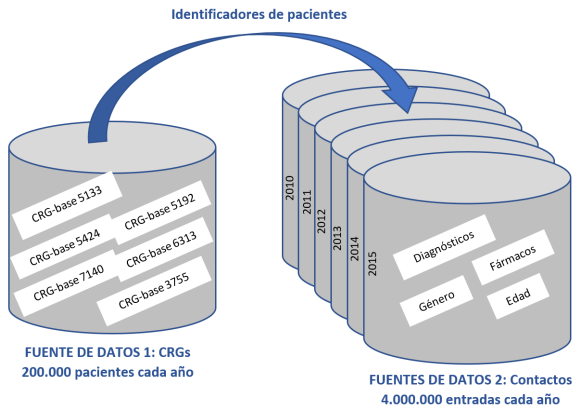


Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 Análisis exploratorio de datos
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple
- 6 Conclusiones y líneas futuras

Bases de datos

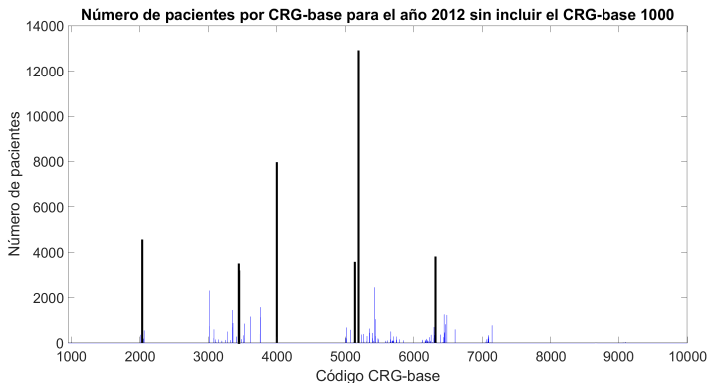
- **Hospital Universitario de Fuenlabrada (HUF)** nos ha proporcionado varias fuentes de datos **anonimizadas**. Los datos con los que se cuenta son de los pacientes adscritos al HUF y por tanto del **Área de Salud** gestionada por el hospital: Fuenlabrada, Humanes y Moraleja de Enmedio.



Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 **Análisis exploratorio de datos**
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple
- 6 Conclusiones y líneas futuras

Número de pacientes por CRG en 2012

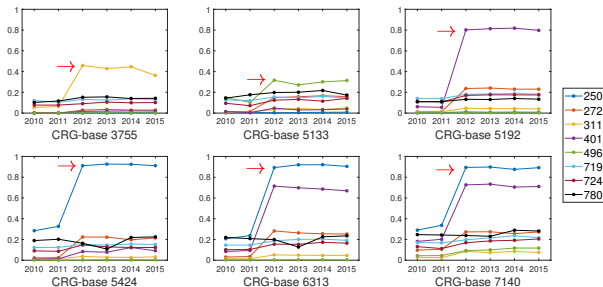


Código CRG-base	Descripción
2030	Una enfermedad aguda significativa
3445	Hiper-lipidemia
3446	Enfermedad tiroidea crónica
4000	Múltiples cronicidades primarias menores

Código CRG-base	Descripción
5138	Asma
5192	HTA
6313	HTA+DM

Evolución temporal: tasa presencia diagnósticos y ATCs

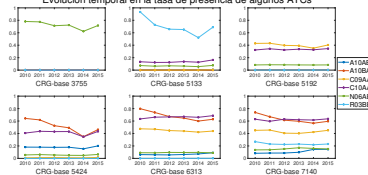
Evolución temporal en la tasa de presencia de algunos diagnósticos



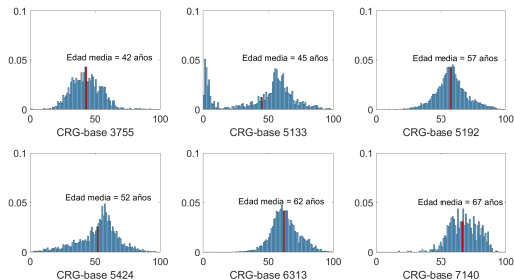
CRG-base	Descripción	Nº pacientes
3755	Depresión	1580
5133	EPOC	964
5192	HTA	12921
5424	DM	2456
6313	HTA+DM	3809
7140	HTA+DM+otra	777

CRG-base	Códigos relacionados
3755	311, N06AB
5133	496, R03BB
5192	401, C09AA, C10AA
5424	250, A10BA, C10AA
6313	250, 401, A10BA, C09AA, C10AA
7140	250, 401, A10BA, C09AA, C10AA

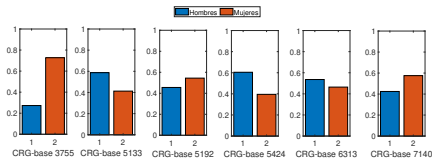
Evolución temporal en la tasa de presencia de algunos ATCs



Distribución de edad y género



Distribución normalizada de edad

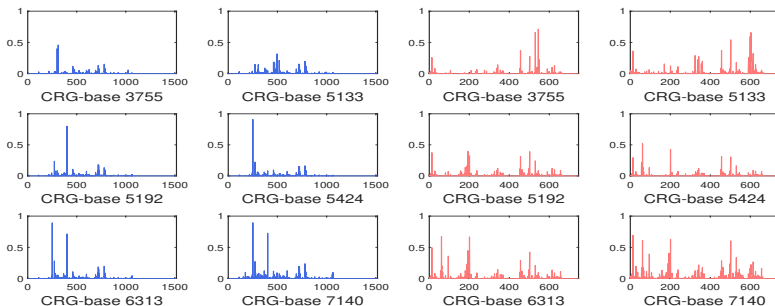


Género

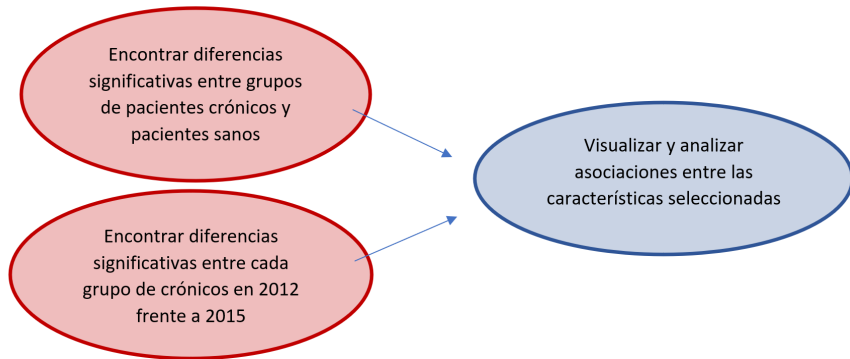
Perfiles de presencia de diagnósticos y códigos ATC

	CRG-base XXXX en 2012							
	Códigos-diagnósticos				Códigos-ATC			
	Código '001'	Código '002'	(...)	Código 'M997'	Código 'A01AA'	Código 'A01AB'	(...)	Código 'V30ZZ'
Paciente 1	0	0	(...)	0	1	0	(...)	1
Paciente 2	1	0	(...)	1	1	1	(...)	0
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
Paciente N	1	0	(...)	1	1	1	(...)	0

Perfil presencia del CRG-base XXXX	$\frac{\sum \text{Código '001'}}{N}$	$\frac{\sum \text{Código '002'}}{N}$	(...)	$\frac{\sum \text{Código 'M997'}}{N}$	$\frac{\sum \text{Código 'A01AA'}}{N}$	$\frac{\sum \text{Código 'A01AB'}}{N}$	(...)	$\frac{\sum \text{Código 'V30ZZ'}}{N}$
---------------------------------------	--------------------------------------	--------------------------------------	-------	---------------------------------------	--	--	-------	--



Objetivos



Test proporciones + remuestreo *bootstrap*

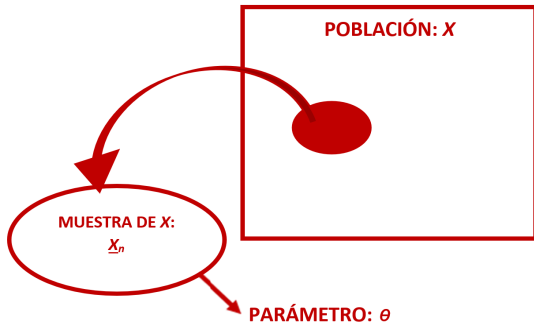
Análisis de Correspondencias Múltiple

Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 Análisis exploratorio de datos
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple
- 6 Conclusiones y líneas futuras

Fundamentos teóricos: contraste de hipótesis

- Un contraste o test de hipótesis es un procedimiento para contrastar la validez de una **hipótesis nula** (H_0) frente a una **hipótesis alternativa** (H_1), sobre un parámetro θ de una población X , a partir de un conjunto de n muestras (\underline{X}_n) de dicha población.



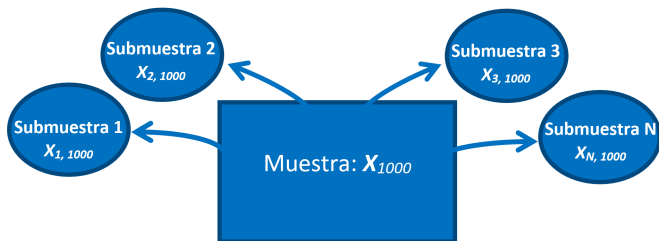
- Cuando el parámetro que se desea conocer es la diferencia de proporciones, se denomina **test de proporciones**.

Fundamentos teóricos: intervalos de confianza y *bootstrap*

- Para llevar a cabo el contraste, inicialmente, se tiene solo una muestra de cada grupo de población estudiado. Por tanto, al realizar el test para cada pareja de CRGs-base en busca de aumentos o disminuciones en la presencia de un determinado código (i), **sólo se dispone de un valor del estadístico de contraste:**

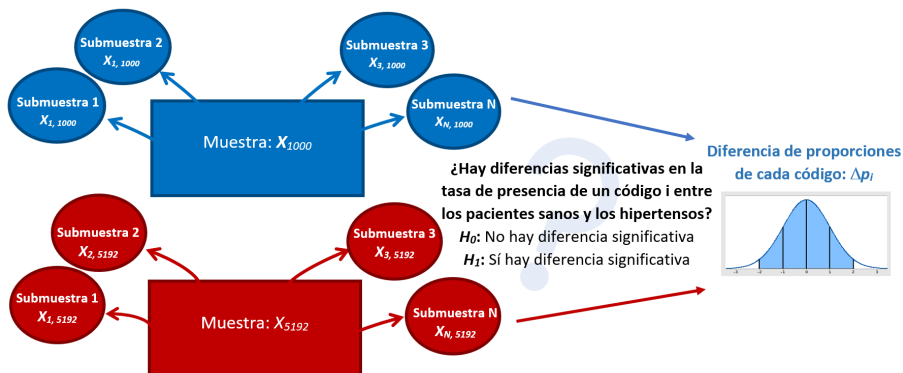
$$\Delta p_i = p_{1000,i} - p_{5192,i}$$

- Para obtener una distribución de Δp_i de cada código, se hará uso del método de **remuestreo bootstrap**. Una vez obtenida la distribución se pueden hallar **intervalos de confianza**.



Fundamentos teóricos: test proporciones + *bootstrap*

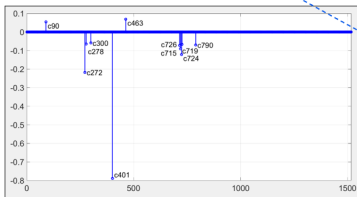
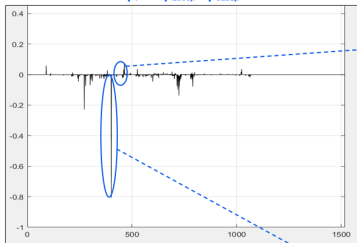
- En este TFG se realiza el siguiente test de proporciones:



Fundamentos teóricos: representaciones gráficas

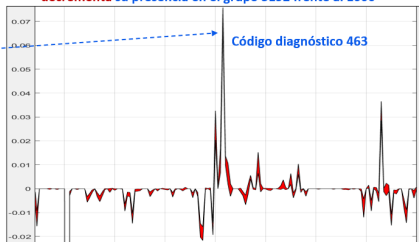
Intervalos de confianza de:

$$\Delta p_i = p_{1000,i} - p_{5192,i}$$

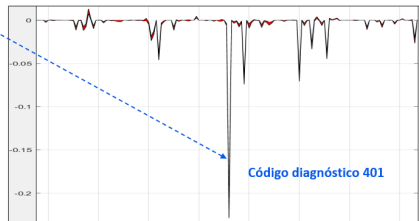


Límite inferior/superior del intervalo de confianza de los códigos seleccionados

Intervalo de confianza ampliado de un código que decrementa su presencia en el grupo 5192 frente al 1000



Código diagnóstico 463



Código diagnóstico 401

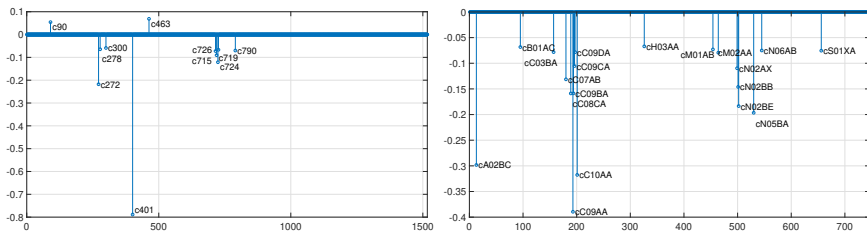
Intervalo de confianza ampliado de un código que incrementa su presencia en el grupo 5192 frente al 1000

Evolución del estado de salud

- Algunas **evidencias** permiten validar el método empleado:
 - **Depresión** → 300 (trastorno depresivo).
 - **EPOC** → 496 (obstrucción crónicas de las vías respiratorias).
 - **HTA** → 401 (hipertensión esencial).
 - **DM** → 250 (diabetes mellitus).
- Adicionalmente se puede concluir:
 - **HTA y DM** → problemas de espalda y articulaciones.
 - **HTA** → problemas de peso y ansiedad.
 - **DM** → hipertensión esencial y cataratas.
 - **HTA, DM y otra cronicidad** → problemas respiratorios, trastornos depresivos y neoplasias malignas.

Evolución del estado de salud

- Se obtienen un total de **76 códigos** (28 diagnósticos y 48 ATCs) con **diferencias significativas** en su tasa de presencia que caracterizarían a los grupos de pacientes crónicos. Estas características serán **seleccionadas** para su análisis mediante el **Análisis de Correspondencias Múltiple**.
- A modo de ejemplo los diagnósticos y ATCs seleccionados para el caso: CRG-base 1000 vs CRG-base 7140:



Límite inferior/superior de Δp_i para el caso: CRG-base 1000 vs CRG-base 7140

Evolución temporal de cada estado de salud

- A través del análisis de la evolución temporal de cada CRG-base, se observan de manera generalizada estos descensos y aumentos:

Aumentos	
A10BD	Hipoglucemiantes orales en asociación
A10AE	Insulinas y análogos acción lenta para inyección
J01CR	Asociaciones de penicilinas, incluidas inhibidores betalactamasa
C09AA	Inhibidores enzima convertasa angiotensina solos

Códigos que aumentan su presencia de manera generalizada

- Estos aumentos podrían indicar una **buena adherencia** a estos tratamientos.

Descensos	
R05CB	Mucolíticos
M02AA	Antiinflamatorios no esteroideos para uso tópico
S01XA	Otros oftalmológicos
M01AB	Derivados del ácido acético y sustancias relacionadas

Códigos que disminuyen su presencia de manera generalizada

- Los descensos podrían estar justificados por la **no financiación de los mismos** a través del Real Decreto-ley 16/2012.

Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 Análisis exploratorio de datos
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple**
- 6 Conclusiones y líneas futuras

Fundamentos teóricos: Matriz binaria Z y perfiles

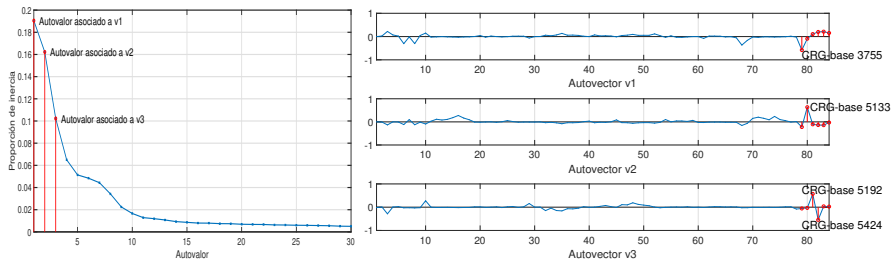
- **Objetivo:** representar gráficamente los datos contenidos en una tabla de tal manera que seamos capaces de visualizar asociaciones entre variables.
- Más variables (categóricas o binarias) \rightarrow más dimensiones \rightarrow reducción de la dimensionalidad \rightarrow pérdida de información \rightarrow inercia.
- Datos \rightarrow **matriz binaria Z** \rightarrow **matriz de Burt B** : $B = Z Z^T$

		Variable 1: Diagnósticos (CIE-9)			Variable 2: Fármacos (ATCs)			Variable 3: CRGs-base		
		Categoría 1: Diagnóstico 001	Categoría 2: Diagnóstico 002	(...)	Categoría 1: Código A01AA	Categoría 2: Código A01AB	(...)	Categoría 1: CRG-base 3755	Categoría 2: CRG-base 5133	(...)
Pacientes	Id 1	0	1	(...)	0	1	(...)	1	0	(...)
	Id 2	0	0	(...)	1	0	(...)	0	1	(...)
	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)

Matriz Binaria Z , perfiles-fila (individuos) y perfiles-columna (categorías)

Fundamentos teóricos: Análisis de Correspondencias Múltiple (ACM)

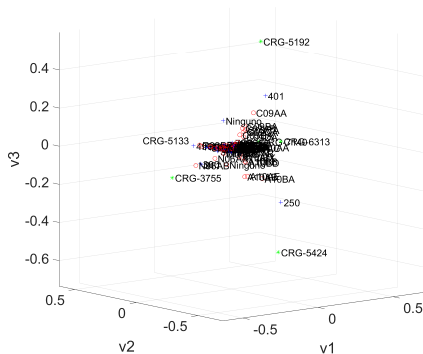
- Descomposición en Valores Singulares (DVS) para crear el subespacio usando los tres autovectores de mayor inercia: $\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$, donde \mathbf{U} y \mathbf{V} son matrices de vectores singulares y los elementos de la diagonal de \mathbf{D}_α son los valores singulares.



- El autovector v1 está discriminando al pacientes con depresión frente al resto, v2 a los pacientes con EPOC y v3 a hipertensos y diabéticos frente al resto.

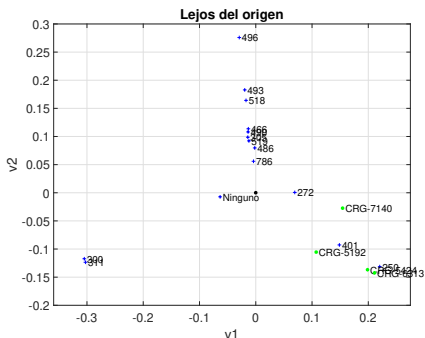
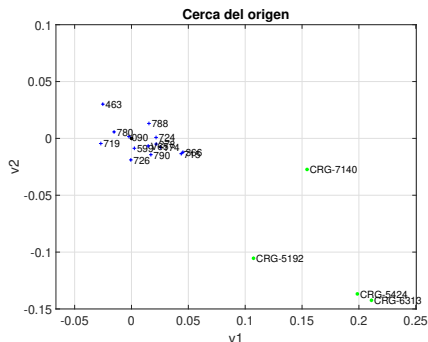
Evolución del estado de salud: asociación entre categorías

- Asociación entre estados de salud (CRGs-base), diagnósticos (CIE-9) y fármacos (ATCs). En total 84 características: 28 diagnósticos + 48 ATCs + 6 CRGs-base. Se ha empleado **balanceo** para tener el mismo número de pacientes en cada grupo.



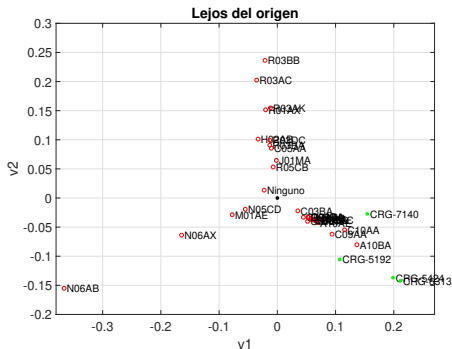
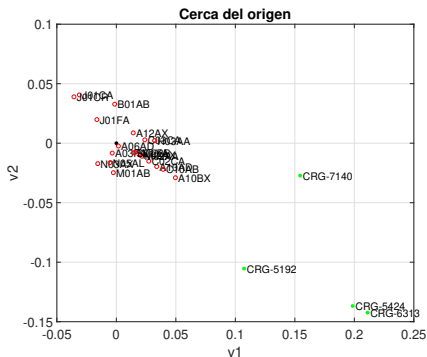
Código CRG-base	Descripción	Código CRG-base	Descripción
3755	Depresión	5424	DM
5133	EPOC	6313	HTA+DM
5192	HTA	7140	HTA+DM+otra cronicidad

Evolución del estado de salud: CRGs-base y diagnósticos



Asociación entre diagnósticos y CRGs-base proyectando sobre v1 y v2.

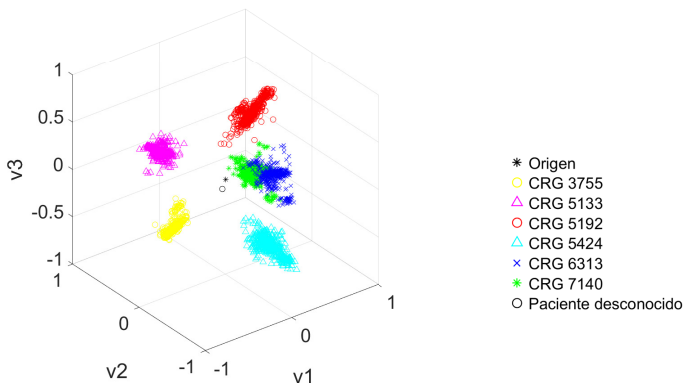
Evolución del estado de salud: CRGs-base y ATCs



Asociación entre ATCs y CRGs-base proyectando sobre v1 y v2.

Evolución del estado de salud: asociación entre pacientes

- Asociación entre los individuos de cada estado de salud (con balanceo) y un paciente desconocido que recoge casos en los que un código no lo presenta ningún paciente.



Individuos proyectados en el subespacio creado por v_1 , v_2 y v_3 .

Contenido

- 1 Introducción: contexto, objetivos y conceptos previos
- 2 Bases de datos
- 3 Análisis exploratorio de datos
- 4 Test de proporciones y remuestreo *bootstrap*
- 5 Análisis de Correspondencias Múltiple
- 6 Conclusiones y líneas futuras**

Conclusiones y líneas futuras

- Conclusiones:

- Pacientes con **HTA y DM**: tienen asociados **problemas de espalda, de articulaciones y trastornos alimentarios**.
- Pacientes con **DM**: diagnósticos relacionados con **HTA y cataratas**.
- Pacientes con **HTA**: tienen asociados **trastornos de ansiedad**.
- **Descenso** en la dispensación de algunos fármacos → Real Decreto-ley 16/2012.
- **Aumento** en la dispensación de algunos fármacos → mejor adherencia al tratamiento.

- Líneas futuras:

- **Aumentar la precisión** de los códigos CRGs, diagnósticos y ATCs.
- **Dividir** cada grupo de pacientes por **rangos de edad**.
- Análisis en **términos de ocurrencia**.
- Diseño de un **modelo predictivo**.

¡MUCHAS GRACIAS!