

Máster Interuniversitario en Estadística e Investigación Operativa UPC-UB

Título: Análisis del perfil de citoquinas, adipoquinas e inmunoglobulinas en la leche materna en los primeros días de lactancia

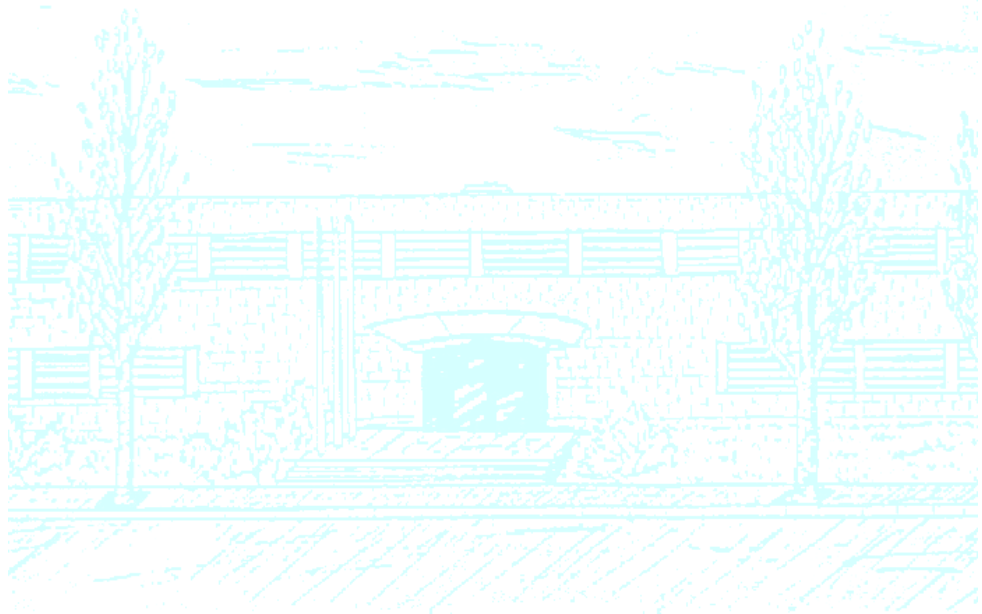
Autor: Aina Fernández Bargalló

Director: Antonio Miñarro Alonso y Esteban Vegas Lozano

Departamento: Dpto. de Genética, Microbiología y Estadística

Universidad: Universidad de Barcelona

Convocatoria: Ordinaria (junio)



Resumen

La leche materna es la ingesta básica que tiene el bebé para su alimentación y defensa. Es por eso por lo que sus componentes son un tema en continuo estudio; a pesar que ya se han detectado muchos de ellos, se sabe que hay muchos más componentes por descubrir. Dicha composición va cambiando desde que el bebé nace hasta que la leche se puede considerar madura, durante lo que se definen las tres etapas de la leche materna, y es muy importante conocer los detalles en la evolución de su composición. En el caso de este estudio nos centramos en la segunda etapa, la leche de transición, para conocer los cambios en la composición de citoquinas, inmunoglobulinas y adipocinas en la leche materna en los primeros días de lactancia además de los posibles factores maternos que pueden influir en estos cambios, a partir de los valores inmunológicos y algunas características personales de 75 pares de madres e hijos.

Palabras clave: *leche materna, valores inmunológicos, preprocesamiento, análisis estadístico*

Clasificación AMS: 62P10

Índice

Introducción	1
I. Marco teórico: La leche materna y su composición	2
1. Inmunoglobulinas, citoquinas y adipoquinas	2
2. Composición de la leche materna	3
3. Fases de la leche materna	4
3.1. El calostro	5
3.2. Leche de transición	5
3.3. La leche madura	6
II. Diseño y objetivos del experimento	7
1. Diseño del experimento	7
2. Objetivos del estudio	8
2.1. Objetivo 1: Comparación entre los dos tiempos	8
2.2. Objetivo 2: Factores influyentes en la composición de la leche	8
III. Metodología estadística	9
1. Estadística descriptiva	9
2. Análisis univariante	10
2.1. Prueba de la T de Student	12
2.2. Modelos de Tobit	13
2.3. Prueba de la Chi Cuadrado	14
2.4. Prueba de Fisher	15
2.5. Suma de rangos de Wilcoxon	15
2.6. Prueba de rangos y signos de Wilcoxon	16
3. Análisis multivariante	17
3.1. Análisis de Factores Múltiples	17
3.2. Random Forest	18
IV. Resultados	20
Objetivo 1. Comparación entre los dos tiempos	20
Análisis descriptivo inicial y transformación de los datos	20
Comparación entre los dos tiempos	20
Análisis de Factores Múltiples	21
Objetivo 2. Factores influyentes en la composición de la leche.	26
Análisis descriptivo inicial y transformación de los datos	26
¿Es la dieta un factor influyente?	26
Caracterización de los dos grupos de madres	27

V. Presentación de resultados: Dashboard	33
Conclusiones	35
Referencias	36
Anexos	38
Anexo A. Resultados del Objetivo 1	38
A.1. Análisis descriptivo inicial	38
A.2. Comparación entre los dos tiempos	43
A.3. MFA para los dos tiempos	45
A.4. MFA para el incremento entre tiempos	48
Anexo B. Resultados del Objetivo 2	49
B.1. Análisis descriptivo inicial	49
B.2. Estudio longitudinal: peso, BMI e infecciones	51
B.3. Comparación entre las dos dietas	52
B.4. Gráficos de individuos del MFA para ver la relación de los inmunotipos con la dieta y las infecciones	56
B.5. Comparación entre los dos grupos de madres (con outliers)	58
B.6. Comparación entre los dos grupos de madres (sin outliers)	62

Índice de figuras

1.	Partes que componen un boxplot	10
2.	Gráfico de los individuos en la primera y segunda dimensiones (MFA de dos tiempos)	22
3.	Gráfico de los individuos en la primera y segunda dimensiones en el día 7 (MFA de dos tiempos)	23
4.	Representación parcial de los individuos en la primera y segunda dimensiones (MFA de dos tiempos)	24
5.	Medidas para el árbol con madres outliers	31
6.	Medidas para el árbol sin madres outliers	31
7.	Captura de pantalla de una pestaña del dashboard del primer objetivo	33
8.	Captura de pantalla de una pestaña del dashboard del segundo objetivo	34
9.	Boxplots múltiples: Comparación entre antes y después de la transformación	43
10.	MFA para los dos tiempos: Porcentajes de variabilidad explicados por cada dimensión	45
11.	MFA para el incremento entre tiempos: Porcentajes de variabilidad explicados por cada dimensión	48
12.	MFA para el incremento entre tiempos: Gráfico de los individuos (1a y 2a dimensiones)	48
13.	Boxplots múltiples: Comparación entre antes y después de la transformación (variables numéricas)	50
14.	Boxplots de la evolución del índice de masa corporal a lo largo de los meses	51
15.	Boxplots del peso a lo largo de los meses	51
16.	Boxplots de las infecciones a lo largo de los meses	52
17.	Relación de la dieta con los inmunotipos en el gráfico de individuos	56
18.	Relación de las infecciones de los 0 a los 6 meses después del parto con los inmunotipos en el gráfico de individuos	56
19.	Relación de las infecciones de los 0 a los 12 meses después del parto con los inmunotipos en el gráfico de individuos	57
20.	Relación de las infecciones de los 0 a los 24 meses después del parto con los inmunotipos en el gráfico de individuos	57

Índice de tablas

1.	Variables con diferencias entre el día 7 y el día 15	21
2.	Distribución de las madres según los dos inmunotipos	27
3.	Variables con diferencias entre los grupos de madres (con outliers)	28
4.	Variables con diferencias entre los grupos de madres (sin outliers)	29
5.	Porcentaje de ceros por variable	38
6.	Estadísticos iniciales (immunoglobulinas, citoquinas y adipoquinas)	39
7.	Estadísticos iniciales (variables calculadas)	40
8.	Estadísticos después de la transformación (immunoglobulinas, citoquina y adipoquinas)	41
9.	Estadísticos después de la transformación (variables calculadas)	42
10.	Resultados de las pruebas para la comparación entre los dos tiempos (immunoglobulinas)	43
11.	Resultados de las pruebas para la comparación entre los dos tiempos (citoquinas)	44
12.	Resultados de las pruebas para la comparación entre los dos tiempos (adipoquinas)	44
13.	Resultados de las pruebas para la comparación entre los dos tiempos (variables calculadas)	44
14.	MFA para los dos tiempos: Contribución de cada variable en cada una de las dimensiones	46
15.	MFA para los dos tiempos: Correlación de cada variable con cada una de las dimensiones	47
16.	Estadísticos iniciales (variables numéricas)	49
17.	Estadísticos después de la transformación (variables numéricas)	50
18.	Resultados de las pruebas para la comparación entre las dos dietas (immunoglobulinas)	52
19.	Resultados de las pruebas para la comparación entre las dos dietas (citoquinas)	53
20.	Resultados de las pruebas para la comparación entre las dos dietas (variables calculadas)	54
21.	Resultados de las pruebas para la comparación entre las dos dietas (variables numéricas)	55
22.	Resultados de las pruebas para la comparación entre las dos dietas (contadores)	55
23.	Resultados de las pruebas para la comparación entre las dos dietas (variables categóricas)	55
24.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (immunoglobulinas)	58
25.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (citoquinas)	59
26.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (adipoquinas)	59
27.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (variables calculadas)	60
28.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (variables numéricas)	61
29.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (contadores)	61
30.	Resultados de las pruebas para la comparación entre los grupos de madres con outliers (variables categóricas)	61

31.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (inmunoglobulinas)	62
32.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (citoquinas)	63
33.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (adipoquinas)	63
34.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (variables calculadas)	64
35.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (variables numéricas)	65
36.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (contadores)	65
37.	Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (variables categóricas)	65

Introducción

La composición de la leche materna es un campo en continuo estudio; a pesar de haberse descubierto ya muchos de sus componentes, se sabe que aún quedan muchos que siguen sin conocerse. Estos componentes van cambiando a medida que pasa el tiempo, desde que nace el bebé hasta llegar a lo que se considera la "leche madura".

En este estudio nos vamos a enfocar en lo que se conoce como la leche de transición, que empieza a generarse a partir del séptimo día después del parto. En esta etapa hay valores inmunológicos muy presentes, como la inmunoglobulina IgA, que se encuentra en grandes concentraciones y ha sido la más estudiada a lo largo de los años. No obstante, hay otros valores, como las citoquinas o las adipocinas, que se encuentran en menores concentraciones, y no suelen tenerse en cuenta en estudios como el que hemos realizado. Es por eso por lo que en este caso, se ha decidido incluirlos.

Este estudio tiene dos objetivos principales, que se explicarán con más detenimiento en apartados posteriores. Por un lado, se quiere comprobar si hay diferencias en la composición de la leche de transición en dos momentos del tiempo distintos: el séptimo día después del parto, y el decimoquinto día después del parto. Por otro lado, se intentará determinar qué factores influyen en la composición de la leche materna. Para ello, hemos realizado este proyecto en colaboración con el departamento de Bioquímica y Fisiología de la facultad de Farmacia y Ciencias de los Alimentos (Universidad de Barcelona), que nos ha proporcionado los datos y los conocimientos necesarios para llevar a cabo el estudio, con el instituto de Investigación en Nutrición y Seguridad Alimentaria (INSA-UB) y con el Instituto de Agroquímica y Tecnología de los Alimentos (IATA-CSIC).

Mi interés por el área de la bioestadística y la atractiva oportunidad de colaboración en un estudio con datos reales, cuyos resultados van a ser relevantes para el área de estudio, ha sido lo que me ha motivado a escoger este estudio como Trabajo de Fin de Máster.

Este trabajo se divide en cuatro grandes partes. En la primera parte, se ha querido hacer una breve explicación del marco teórico del trabajo, desde una introducción a aquellos valores inmunológicos que se van a contemplar en el estudio, hasta una descripción sobre la composición de la leche materna y sobre cada una de sus fases.

En una segunda parte, se ha descrito cómo se realizó el diseño del experimento, y se ha hecho una explicación más extensa de los dos objetivos principales del estudio.

La cuarta parte está dedicada a aquellas metodologías estadísticas que se han usado durante el estudio, haciendo una descripción de cada una de ellas y añadiendo una breve explicación de cómo se puede realizar cada una de las técnicas usando el software estadístico R.

En la quinta parte, se presentan los resultados obtenidos en cada uno de los objetivos, después de aplicar las técnicas convenientes de las descritas en el apartado anterior.

Además, se incluye una breve explicación de cómo se han presentado los resultados a los investigadores durante las reuniones que hemos tenido al largo de este estudio para hacer un seguimiento del trabajo realizado e incluir nuevas inquietudes que pudiesen surgir a medida que éste avanzaba.

Todo este trabajo se ha realizado mediante el software estadístico R.

I. Marco teórico: La leche materna y su composición

Antes de entrar a la parte estadística del trabajo, conviene hacer una breve explicación de los datos y variables que se van a estudiar en este estudio.

En este estudio se van a estudiar diferentes variables que componen la leche materna, así como ciertas características de las madres y sus hijos durante la etapa de lactancia, para comprobar si la composición de la leche en la etapa de transición varía entre períodos cortos de tiempo y si hay factores que afecten a dicha composición.

Debemos saber, antes que nada, que la lactancia materna es *el proceso por el que la madre alimenta a su hijo recién nacido a través de sus senos, que segregan leche inmediatamente después del parto, que debería ser el principal alimento del bebé al menos hasta los dos años*[21].

La leche materna humana se considera la nutrición ideal para los recién nacidos, ya que no solo contiene los nutrientes necesarios para su crecimiento y desarrollo, sino que también numerosos factores bioactivos que contribuyen a los efectos beneficiosos de la lactancia materna. Todavía no se han descubierto la totalidad de estos factores, y de hecho la lista cada vez es más larga, pero incluye hormonas, factores de crecimiento y citoquinas, entre otros [9]. De hecho, la alimentación exclusiva con leche materna durante los primeros 6 meses de vida, con la lactancia materna continua durante 1 a 2 años de vida o más, se reconoce como el estándar normativo para la alimentación infantil [6].

Una vez definido en qué consiste la lactancia materna, conviene saber qué etapas tiene y de qué se compone la leche materna. De esta forma tendremos un marco teórico que nos ayudará a comprender mejor los resultados estadísticos obtenidos en el estudio.

Se cree que un bebé nace sin microbios en su tracto intestinal, pero dentro de las 24 horas posteriores al nacimiento, éste está poblado. Durante esta transición la leche materna juega un papel protector, y continúa la exposición del bebé al sistema inmunológico de la madre que empezó antes del nacimiento [1].

La composición de la leche que se segrega cambia notablemente durante el proceso de lactancia, sobretodo durante las dos primeras semanas de vida del bebé. De hecho, la leche materna se considera un líquido inteligente, ya que se adapta a las necesidades del bebé en función de sus necesidades nutricionales y su etapa de crecimiento. Es por eso por lo que la Organización Mundial de la Salud recomienda hacer una lactancia exclusiva de leche materna hasta los seis meses, y luego combinarla con alimento hasta los dos años.

A continuación se van a definir aquellos valores inmunológicos más relevantes para el estudio, y seguidamente se va a hacer una explicación de la composición de la leche en general, y seguidamente de dicha composición en cada una de las fases que conforman este proceso de lactancia.

1. Inmunoglobulinas, citoquinas y adipocinas

En este estudio se ha querido estudiar la concentración de las inmunoglobulinas y de algunas citoquinas y adipocinas en la composición de la leche durante la etapa de transición de la leche, que se definirá más adelante. Por lo tanto es conveniente hacer una breve explicación de dichos grupos de compuestos.

Antes de eso, conviene decir que, en la literatura, los valores más estudiados son los de las inmunoglobulinas, en especial la inmunoglobulina IgA. Aun así, en este estudio se han querido incluir las citoquinas y las adipocinas, a pesar de estar presentes en concentraciones muy bajas.

La información sobre los niveles maternos de marcadores inmunitarios a los que puede estar expuesta la descendencia se obtiene mediante análisis de sangre que se realizan durante el embarazo. Estos marcadores,

tanto los obtenidos en la sangre materna como los del líquido amniótico o los de la sangre del cordón umbilical, son necesarios para comprender la exposición de la descendencia, ya que por ejemplo las madres atópicas pueden producir una cantidad diferente de citoquinas que las madres que no lo son [1].

Por un lado, las inmunoglobulinas son *proteínas plasmáticas sintetizadas por los linfocitos B maduros y las células plasmáticas, en respuesta a la estimulación por un antígeno, y que actúan como anticuerpos, para la defensa específica del organismo* [19].

Hay distintos tipos de inmunoglobulinas. La más presente en la leche materna es la IgA, que es el *compuesto inmunoquímico que predomina en la leche humana; forma parte de la primera línea de defensa de la mucosa intestinal contra gérmenes capaces de invadir el epitelio. La principal función de la IgA es bloquear la adherencia de agentes patógenos a la mucosa intestinal* [18]. La IgA es un anticuerpo fundamental para la inmunidad de las mucosas, donde se detecta más concentración de este anticuerpo que de todos los demás combinados. Además, es muy resistente a la degradación, por lo que puede funcionar en los entornos hostiles de los tractos digestivo y respiratorio [1]. Esta inmunoglobulina se puede dividir en dos clases: la IgA1 y la IgA2; la segunda es la predominante en la leche.

Además, la IgA actúa uniéndose directamente a antígenos microbianos específicos, de forma que bloquea la adhesión y modula la función inmune local, favoreciendo la colonización microbiana normal del intestino [9].

Aun así, también se incluyen en el estudio otras inmunoglobulinas, como las siguientes: IgG (que a su vez se divide entre IgG1, IgG2, IgG3 y IgG4), IgE y IgM. Por un lado, la IgG es el anticuerpo más común en el sistema circulatorio. De entre sus subclases, la IgG4 se encuentra en bajas concentraciones, y de hecho es la menos abundante en la leche. Por otro lado, la IgM es el anticuerpo más grande en el sistema circulatorio humano y se considera que es el primer anticuerpo producido en una respuesta primaria de anticuerpos [1].

Las citoquinas, por su parte, son *pequeñas proteínas que son cruciales para controlar el crecimiento y la actividad de otras células del sistema inmunitario y las células sanguíneas*[11]. En el caso de este estudio, se cuenta con un gran número de citoquinas distintas, entre las que se encuentran la IL27, la IL1b, la IL2, la IL4 o la IL5, entre muchas otras.

Las citoquinas, como ya se ha comentado, normalmente no se añaden en los estudios sobre los valores inmunológicos de la leche materna por su baja concentración. De hecho, la mayoría de ellas no se encuentran en las fórmulas comerciales de leche para bebés, lo que indica que no se consideran imprescindibles, aunque no hay suficientes estudios como para confirmarlo [15].

Por último, las adipoquinas son *una serie de moléculas de señalización celular secretadas por el tejido adiposo*. En este estudio tan solo se van a estudiar dos componentes que formen parte de este grupo: las leptinas, unas hormonas que regulan el apetito, y las adiponectinas, unas hormonas que ayudan a quemar grasa. Además, se ha visto que pueden tener un peso importante en el desarrollo del bebé gracias a su presencia en la leche materna.

Adicionalmente, en el estudio se cuentan con algunas variables que se han creado a partir de los valores que toman las anteriores. Durante todo el estudio nos vamos a referir a ellas como "variables calculadas". Algunos ejemplos de ellas son la Th1, que se calcula tanto para las inmunoglobulinas (siendo la suma de la IgG1, la IgG2 y la IgG3) como para las citoquinas (siendo en este caso la suma de la IL27, la IL1 β , la IL2, la IL12, la Ifn- γ y la TFN- α); el Ratio L/A, que es la división entre la medida de las leptinas y la de las adiponectinas; o las variables que se corresponden a cada una de las inmunoglobulinas en porcentaje.

2. Composición de la leche materna

Como hemos comentado anteriormente, a menudo se refiere a la leche materna como un alimento inteligente. Al ser el primer alimento del bebé, la leche materna contiene todos los nutrientes básicos esenciales, como

son los carbohidratos, las proteínas (que ayudan al crecimiento y desarrollo de tu bebé, activan su sistema inmunitario y desarrollan y protegen las neuronas cerebrales) y las grasas. Además, contiene la suficiente cantidad de agua para mantenerle hidratado.

Aun así, la leche materna no aporta tan solo nutrición, sino que aporta otros componentes extra que no se encuentran en las leches de fórmula. No obstante, la composición de la mezcla de compuestos de la leche que interactúan no solo difiere entre las mujeres sino que también puede influir en su composición el grado de prematuridad, y puede variar durante el período de lactancia, desde el calostro, con las concentraciones más altas, pasando por la leche de transición a la madura, como se verá en el apartado siguiente [9].

Los componentes nutricionales de la leche materna se derivan de tres fuentes: algunos de los nutrientes de la leche se originan por síntesis en el lactocito, algunos son de origen dietético y algunos se originan en las reservas de la madre. En general, la calidad nutricional de la leche materna se conserva en gran medida, pero la atención a la dieta materna es importante para algunas vitaminas y la composición de ácidos grasos de la leche materna [6].

La leche materna contiene componentes nutricionales, entre los que se encuentran algunos macronutrientes y algunos micronutrientes, componentes bioactivos, factores de crecimiento y factores inmunológicos. A continuación, se van a detallar algunos de los componentes más relevantes.

Por un lado, la leche materna contiene millones de células vivas, como glóbulos blancos ¹, que refuerzan el sistema inmunitario, o citoblastos ², que ayudan al desarrollo y curación de los órganos. Además, contiene cientos o miles de moléculas bioactivas distintas que protegen contra la infección y la inflamación y contribuyen a la maduración inmunológica, el desarrollo de órganos y la colonización microbiana saludable [6].

También contiene muchos azúcares complejos (oligosacáridos) que actúan como prebióticos y evitan que las infecciones lleguen al torrente sanguíneo, además de reducir el riesgo de inflamación cerebral, y enzimas, que aceleran las reacciones químicas del cuerpo y ayudan a los sistemas digestivo e inmunitario de tu bebé y permiten la absorción del hierro.

Además de estos componentes, en la leche materna también se pueden encontrar factores de crecimiento, que favorecen un desarrollo saludable; hormonas, envían mensajes entre los tejidos y órganos para garantizar su correcto funcionamiento; vitaminas y minerales, que favorecen el crecimiento saludable y el funcionamiento de los órganos, y además ayudan a la formación de los dientes y huesos del bebé; ácidos grasos de cadena larga, que tienen un papel fundamental en el desarrollo del sistema nervioso del bebé y contribuyen al desarrollo saludable del cerebro y los ojos; y diferentes moléculas de microARN, que tienen la función de regular la expresión genética y de ayudar a evitar o detener el desarrollo de enfermedades, reforzar el sistema inmunitario del bebé y participar en la remodelación del pecho.

Aun así, la leche materna también tiene otro componente principal que es el que más nos ocupa en este estudio: los anticuerpos, también conocidos como inmunoglobulinas. Éstos se encargan principalmente de proteger al bebé contra las enfermedades e infecciones, ya que neutralizan las bacterias y los virus.

Además de estos componentes, hay muchos otros y cada vez se van descubriendo de nuevos. Como ya se ha comentado, dichos componentes van cambiando con el paso del tiempo, por lo que vamos a detallar cada una de las fases de la leche materna y sus componentes en cada una de ellas.

3. Fases de la leche materna

En concreto, se definen tres fases en la composición de la leche materna: el calostro (o primera leche), la leche de transición, que se produce entre la segunda y la tercera semana, y la leche madura, que es la que se produce a partir de la cuarta semana.

¹Células encargadas de defender al organismo de las infecciones y ayudar a eliminar los residuos y desechos de los tejidos.

²Núcleo protoplasmático de una célula, como lugar desde el que se desarrolla la célula.

Los ingredientes básicos de la leche son los mismos en todas las fases mientras dure la lactancia, pero sus niveles subirán o bajará en función de las circunstancias.

A pesar de que en este estudio, como ya hemos comentado, nos vamos a centrar en la leche de transición (de hecho se tomaron las muestras el día 7 y el día 15 de la lactancia) es relevante hacer una breve explicación de cada una de las fases.

3.1. El calostro

La leche que se obtiene tras el parto se denomina calostro. Ésta es muy importante a la hora de nutrir y proteger al bebé, ya que acaba de nacer y es muy vulnerable. Esta leche se produce en cantidades pequeñas los primeros días después del parto.

El calostro es rico en componentes inmunológicos como la IgA secretora, la lactoferrina o los leucocitos, así como factores de desarrollo. Tiene concentraciones relativamente bajas de lactosa, por lo que sus funciones primarias son más inmunológicas que nutricionales. Además los niveles de sodio, cloruro y magnesio son más altos que en etapas posteriores, al contrario que los de potasio y calcio, que son más bajos [6].

Durante esta etapa, la leche contiene más de 1000 proteínas y aminoácidos, que además de contribuir al crecimiento y desarrollo del bebé, activan su sistema inmunitario y protegen sus neuronas.

En concreto, hay dos clases de proteínas en la leche materna: caseína, que ayuda a mantener al bebé saciado durante más tiempo y tiene propiedades antimicrobianas, y el suero de leche, es rico en anticuerpos y se mantiene líquido, por lo que es más fácil de digerir. Durante el calostro las proporciones de estos dos tipos de leche suelen ser muy desiguales: normalmente se tiene aproximadamente un 90 % de caseína y tan solo un 10 % de suero de leche.

También cuenta con elevados niveles de anticuerpos, glóbulos blancos y citoblastos que refuerzan el sistema inmunitario y contribuyen al desarrollo de los órganos, protegiéndolo de enfermedades e infecciones.

Además, contiene más de 200 oligosacáridos (o lo que antes hemos definido como azúcares complejos) y es muy rica en minerales y vitaminas A, E y K, que ayudan al correcto funcionamiento del organismo y ayudan a la formación de dientes y huesos.

3.2. Leche de transición

Es la leche que se produce durante la segunda y la tercera semana, concretamente entre los días 5 y 14 tras el parto. Durante esta etapa, se produce un aumento de la cantidad de leche muy grande, ya que el bebé empieza a necesitar más alimento y un equilibrio diferente de nutrientes. De hecho, se puede llegar a doblar la cantidad de leche diaria que se produce con respecto a la producción durante el calostro. Aunque comparte algunas de las características del calostro, representa un período de producción de leche acelerada para satisfacer las necesidades nutricionales y de desarrollo del bebé, que está en rápido crecimiento [6].

Aunque durante esta etapa la leche mantiene todas las propiedades del calostro, los niveles de grasa, calorías y azúcares –especialmente de lactosa– son mucho mayores, contribuyendo al desarrollo del sistema nervioso, del cerebro y ojos.

Además, el contenido proteico de la leche materna también sufre cambios. Mientras que las proporciones de los dos tipos de proteínas eran muy desiguales durante el calostro, en la etapa de transición se van acercando más, aunque para que las proporciones sean aproximadamente iguales tendrá que pasar como mínimo un mes de lactancia.

Por otro lado, la concentración de las enzimas protectoras y anticuerpos también se modifica, ya que durante esta fase el bebé ya empieza a desarrollar su propio sistema inmunitario.

En cuanto a los valores inmunológicos, que son los que se estudiarán en este trabajo, tienen mucha presencia en la leche materna durante la primera etapa, y van disminuyendo progresivamente. Durante la fase de transición, la leche tiene mucha concentración de inmunoglobulinas, en especial de la inmunoglobulina IgA, que es la primera línea de defensa frente a la infección: es muy importante para los bebés porque adquieren defensas en las primeras semanas de vida, ya que tienen un sistema inmunitario muy débil.

Aun así, la leche no tiene mucha presencia de citoquinas y adipoquinas, por lo que su concentración toma valores muy bajos, y en ocasiones llega a ser inexistente.

3.3. La leche madura

La leche madura es la leche que se produce a partir de la cuarta semana después del parto. En este momento, la leche es totalmente madura, y presenta menores concentraciones de proteínas, minerales y vitaminas que el calostro pero añade un gran número de componentes bioactivos, como hormonas, inmunoglobulinas y enzimas.

A partir de las cuatro semanas, los nutrientes y los niveles de ingredientes de la leche madura suelen permanecer bastante constantes. No obstante, puede que la composición de la leche materna siga cambiando día a día y toma a toma.

II. Diseño y objetivos del experimento

Una vez se ha explicado el marco teórico sobre el que se basa el estudio, es conveniente definir cómo se ha diseñado el experimento que se ha llevado a cabo para obtener los datos y qué objetivos se plantean en el estudio.

1. Diseño del experimento

El estudio que se ha realizado se basa en un estudio llamado MAMI , una cohorte de nacimiento centrada en la microbiota maternoinfantil durante la vida temprana [17], el objetivo de la cual era caracterizar los microbios maternos transferidos a los recién nacidos durante sus primeros 24 meses de vida e identificar los posibles factores ambientales, de la madre y del hijo (como la edad gestacional o la dieta) que afectan a la colonización microbiana temprana.

Aunque en el caso de este estudio se persigue un objetivo distinto, los datos que se van a usar provienen de dicha cohorte, aunque se dispone tan solo de algunas de las muestras. De hecho, el presente estudio sería un estudio final para terminar el proyecto.

El artículo original es un estudio de cohorte observacional prospectivo longitudinal, lo que quiere decir que aunque el estudio se diseñó y se comienza a realizar en cierto momento del tiempo, los datos se analizan transcurrido un determinado tiempo.

El estudio original incluye los datos de 250 parejas de madres e hijos del área mediterránea de España, que se siguieron durante los primeros dos años de vida del bebé. Aun así, en el presente estudio tan solo se tendrán en cuenta 75 de ellas (se proporcionaron los datos de 105 madres, pero solo 75 de ellas tenían información de las medidas tanto a los 7 días como a los 15 días después del parto).

Para escoger las madres que participarían en el estudio, se hizo un reclutamiento inicial a través de ginecólogos y parteras, así como en centros de atención primaria y hospitales, mediante su difusión con folletos, carteles o fichas informativas, entre otros. Los centros proporcionaban a las mujeres que estaban en la etapa final del embarazo información sobre el estudio, y éstas podían inscribirse en él hasta la semana posterior al nacimiento de su hijo. Además, también se podía encontrar información sobre el estudio en redes sociales.

Dicha reclutación de madres para el estudio empezó en 2015, y se completó el seguimiento de todas ellas a principios de 2019. Los criterios de inclusión requerían que la madre fuera mayor de 18 años, que el embarazo fuera saludable y que se supiese hablar y escribir el español. Además, se excluyeron aquellas madres que no cumplieran alguno de estos criterios o que usaran medicamentos o drogas, padecieran de complicaciones de salud durante el período gestacional o bien alguna enfermedad crónica, o tomaran medicación por alguna patología crónica. Además, todas las madres debían residir en el área Mediterránea para asegurar que todas siguiesen el mismo estilo de vida y exposición ambiental, evitando así la existencia de factores de confusión.

Se analizaron los valores inmunológicos (la concentración de ciertas adipocinas, citoquinas e inmunoglobulinas en la leche materna) durante la fase de transición del embarazo y en distintos momentos del tiempo (a los siete días después del parto y a los quince días).

En la literatura, en los estudios de este tipo se suelen tener en cuenta tan solo las inmunoglobulinas, ya que las citoquinas y las adipocinas aparecen en concentraciones muy bajas. No obstante, en este estudio también se han querido tener en cuenta.

En concreto, se ha recogido información sobre dos adipocinas, dieciocho citoquinas y siete inmunoglobulinas. Las concentraciones de dichas variables se han medido tanto al séptimo día del embarazo como al decimoquinto día. Además, se ha recogido información referente a la dieta que siguen las madres durante el embarazo (diferenciando entre la dieta I y la dieta II), el género del bebé, así como su peso, su índice de masa corporal y el número de infecciones que se le han observado en distintos momentos del tiempo, entre otros.

2. Objetivos del estudio

En el presente estudio se han determinado dos objetivos principales: ver si hay alguna diferencia entre las medidas de los valores inmunológicos tomadas a día siete de las tomadas a día quince, y determinar qué factores influyen en la composición de la leche.

A continuación, se van a detallar cada uno de ellos.

2.1. Objetivo 1: Comparación entre los dos tiempos

El primer objetivo que se propuso en el estudio fue el de observar los cambios en la composición de la leche de día 7 a día 15.

Para ello, se van a tener en cuenta las 75 madres de las que se tiene información en ambos tiempos, de forma que se tendrán dos muestras apareadas, ya que cada variable será medida dos veces sobre la misma madre: una vez en el día 7 del embarazo y la otra a día 15.

Además, tan solo se van a tener en cuenta las variables referentes a los distintos valores inmunológicos: las adipoquinas, las citoquinas y las inmunoglobulinas.

El objetivo de este apartado será el de ver si existe relación entre la composición de la leche a día 7 y a día 15 y, por lo tanto, poder saber si hay diferencias en la composición de la leche entre los dos tiempos del análisis.

2.2. Objetivo 2: Factores influyentes en la composición de la leche

Como segundo objetivo, nos planteamos el investigar si hay algunos factores que influyen en la composición de la leche.

Para ello, se va a usar de nuevo la información de las madres de las que se tomaron las medidas en ambos tiempos. En este objetivo no se van a usar las variables como muestras apareadas, si no que se van a tener en cuenta los dos tiempos como variables independientes, ya que no deseamos compararlos entre ellos.

En este caso, además de utilizar los valores inmunológicos se van a añadir en el estudio aquellas variables relacionadas con las características de la madre y del bebé, como la dieta, el peso o si tuvieron o no infecciones.

Como primer criterio, se va a contemplar la idea de que la dieta puede ser un factor influyente en la composición de la leche. A continuación, se van a explorar el resto de variables para ver cuales pueden influir en la composición de la leche o en la separación de grupos de madres bien diferenciados.

III. Metodología estadística

En este apartado se va a hacer una explicación detallada de aquellas técnicas que posteriormente se aplicarán para dar respuesta a los dos objetivos descritos.

Las técnicas estadísticas que se han usado son las siguientes: la representación gráfica de variables mediante los boxplots, pruebas paramétricas para la significación estadística (en concreto, la prueba de la T de Student, los modelos Tobit y la prueba de la Chi cuadrado), pruebas no paramétricas para la significación estadística (en este caso, la prueba de Wilcoxon) y técnicas de análisis multivariante, en concreto el análisis de factores múltiples y el Random Forest.

Estas técnicas se han aplicado de formas distintas en función del objetivo que se persiguiera en cada caso. A continuación, se hace una explicación detallada para entender en qué consiste cada una de las técnicas, cómo se han aplicado en cada caso y de qué forma se pueden reproducir usando el software estadístico R.

1. Estadística descriptiva

Un primer paso en todo análisis estadístico es el de hacer un análisis descriptivo inicial. Hay que tener en cuenta que este análisis no puede ser igual para las variables numéricas que para las categóricas, ya que de las variables categóricas, a pesar de poder observar el número de registros que hay en cada nivel de la variable, no se puede observar cómo se distribuyen. En cambio, para las variables numéricas sí que se puede observar, y de hecho es importante hacerlo, su distribución.

Para ello, lo más común es calcular algunos estadísticos, como pueden ser la media, la mediana o la desviación estándar, que ayudan a entender el comportamiento y distribución de cada variable. En el caso de este estudio, se han calculado los siguientes estadísticos para cada una de las variables numéricas del estudio: el número de observaciones, la media, la desviación estándar, la mediana, el mínimo, el máximo y el error estándar.

Aun así, también es muy habitual el uso de métodos gráficos. A pesar de que hay muchos métodos, como por ejemplo los histogramas, en este estudio se ha decidido hacer uso de los boxplots.

Los boxplots, o diagramas de cajas, son una forma estandarizada de mostrar la distribución de una base de datos basada en cinco estadísticos: el mínimo, los tres primeros cuartiles y el máximo. Además, da información sobre los outliers, es decir, valores que difieren significativamente del resto. En la figura que se presenta a continuación se puede visualizar su forma y las partes que lo componen.

Como se puede ver en el gráfico, en un boxplot se muestran los valores outliers en forma de bolas, que aparecerán siempre fuera de la caja y el bigote (que es la línea que une el máximo y mínimo con la caja, y que se muestra en azul). El mínimo aparece en el extremo izquierdo del bigote, y se trata del valor observado más pequeño. Yendo hacia la derecha, aparece el primer cuartil (o percentil 25), que es el número medio entre el mínimo y la mediana; así pues, el 25 % de los datos están por debajo de este valor. La mediana, que es la línea amarilla de dentro de la caja, indica el valor que está en la mitad de los datos, es decir, aquel que deja el 50 % de datos por debajo y el 50 % por encima. Cierra la caja el tercer cuartil, o percentil 75, que deja el 75 % de los datos por debajo y es el valor que está en la mitad entre la mediana y el máximo. El máximo aparece en el extremo derecho, y es el valor observado más grande que no se puede determinar como outlier.

Aunque en ocasiones el determinar que un valor es outlier (o atípico) y no un máximo o un mínimo es una cuestión subjetiva, normalmente se considera que es un valor atípico cuando se encuentra a más de 1,5 veces la distancia del rango intercuartílico (que es la amplitud de la caja) del primer o del tercer cuartil (atípico leve), como se muestra en el gráfico, o a 3 veces esa distancia (atípico extremo).

³<https://ichi.pro/es/comprencion-de-los-diagramas-de-caja-103551522880849>

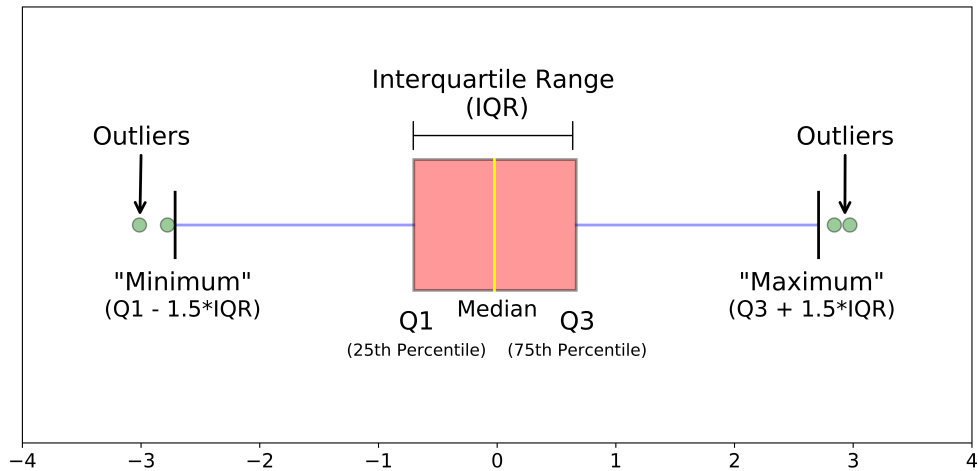


Figura 1: Partes que componen un boxplot ³

Así pues, gracias a los boxplots se puede tener información sobre si los datos están bien distribuidos, es decir, la dispersión de los datos, y sobre los valores atípicos que aparecen, pero también sobre la simetría de los datos, así como de qué tan bien se agrupan y si están sesgados (y, en el caso que lo estén, cómo se sesgan). La dispersión de los datos viene dada por el tamaño de la caja; la simetría, por la posición de la mediana dentro de la caja; y el sesgo, por la largada de las colas (bigotes) de los boxplots.

En el caso de este estudio, los boxplots van a tener dos funciones principales:

- Por un lado, mediante el graficado de los boxplots múltiples por cada grupo de variables (inmunoglobulinas, citoquinas, adipocinas y variables calculadas) se va a visualizar la distribución de los datos, ayudándonos así a determinar si los datos siguen una distribución Normal o si, por contra, es necesaria alguna transformación para normalizarlos;
- Por otro lado, se van a graficar los boxplots de cada una de las variables separándolas por un grupo de interés, que en este caso será cada uno de los tiempos (día 7 y día 15) o cada una de las dietas. De esta forma, se van a poder ver las diferencias en la distribución de una misma variable entre días o entre dietas, por ejemplo. Esto nos servirá de apoyo a las pruebas paramétricas y no paramétricas que se hagan para determinar si hay diferencias entre los dos grupos.

Los boxplots se pueden graficar de forma sencilla en R gracias a la función `boxplot`, especificando la variable que se quiere estudiar y pudiendo cambiar distintos parámetros como los colores o la dirección de los boxplots (horizontal o vertical). En el caso de que se quieran hacer en función de otra variable, para por ejemplo comparar entre tiempos, basta con especificar ambas variables como se haría en un modelo de regresión:

```
boxplot(variable ~ variable grupo)
```

Como ya se ha comentado, no se puede hacer uso de esta técnica para las variables categóricas.

2. Análisis univariante

Las pruebas univariantes se centran en una característica o propiedad de los individuos de estudio. En este trabajo, se usarán tanto pruebas paramétricas como pruebas no paramétricas, que servirán para determinar si hay diferencias entre dos grupos de estudio, como se detallará a continuación.

Por un lado, las pruebas paramétricas son una herramienta estadística que se utiliza para el análisis de los factores de la población. En concreto, en el caso de este estudio servirán para comparar si hay diferencias en las medidas de las diversas variables entre dos grupos, que serán o bien los dos tiempos o bien los dos tipos de dieta.

Este tipo de pruebas se basan en la distribución normal, y por lo tanto, se debe cumplir la condición de normalidad de los datos para que sus conclusiones sean fiables. Además, los datos deben cumplir también con las condiciones de homocedasticidad (es decir, que la variabilidad entre las dos muestras sea uniforme, o dicho de otra forma, que sus varianzas se puedan considerar como iguales) e independencia de los errores (cosa que ocurre cuando los sujetos son asignados de forma aleatoria y se distribuyen de forma normal dentro de cada grupo).

Para este estudio se van a utilizar dos pruebas paramétricas distintas: la prueba de la T de Student y los modelos Tobit (que, a pesar de no ser una prueba como tal, también necesita el cumplimiento de las hipótesis que se han planteado en el párrafo anterior). Adicionalmente, para algunas de las variables se ha realizado la prueba de la χ^2 cuadrado para la comparación de proporciones, debido a que son variables que tan solo toman dos posibles valores (1 si ha sido observado el hecho, y 0 en caso contrario).

Concretamente, se va a usar la prueba T de Student para todas las variables excepto para las que forman parte del grupo de las citoquinas. El motivo de esta decisión es que en las citoquinas hay un número muy elevado de ceros y en estos casos los modelos Tobit son mas adecuados, como se verá a continuación. Además, para aquellas variables que se puedan considerar binarias o contadores (como por ejemplo, si hay o no hay infección) se va a hacer un contraste de proporciones.

La hipótesis nula a contrastar en dichas pruebas es la siguiente: *No hay diferencias entre los dos grupos en la variable*. Por lo tanto, si se obtiene un p-valor significativo en alguna de las variables, se podrá asumir que hay diferencias entre los dos grupos de estudio (ya sean los dos tiempos, o las dos dietas).

Para todas las pruebas paramétricas realizadas, se ha determinado un nivel de confianza del 95 %, por lo que se va a considerar que una variable es estadísticamente significativa cuando su p-valor sea inferior al nivel de significación, igual a 0.05.

Por otro lado, las pruebas no paramétricas (o de distribución libre) son aquellas que se aplican a datos que no tienen una organización normal. La mayoría de las pruebas no paramétricas se basan en la ordenación de las medidas obtenidas, como se verá a continuación.

Aunque en el caso de este estudio se ha aplicado la transformación correspondiente para poder determinar que los datos siguen una distribución normal, se ha hecho la prueba no paramétrica de forma adicional, lo que nos permitirá además comparar los resultados de una prueba paramétrica con los que se obtendrían con la prueba no paramétrica.

En este caso se ha decidido trabajar con las pruebas de Wilcoxon, que son la alternativa a la prueba T de Student para cuando las poblaciones tienen una distribución que no satisface las condiciones necesarias para hacer una prueba paramétrica.

Dichas pruebas trabajan sobre rangos de orden, lo que quiere decir que utiliza las posiciones que ocupan los datos una vez ordenados, por lo que solo es aplicable a variables que tienen valores ordenables.

En el caso de estas pruebas, la hipótesis nula a contrastar dice que *No hay diferencias en las medianas de los dos grupos*. Por lo tanto, la interpretación es la misma que en el caso de las pruebas paramétricas: si una variable es significativa asumiremos que hay diferencias entre los dos grupos de estudio.

En el caso de querer hacer la prueba para muestras independientes, se hará la prueba de la suma de rangos de Wilcoxon; en cambio, si se quiere hacer sobre datos apareados, la opción a escoger es la prueba de rangos y signos de Wilcoxon. Por lo tanto, el procedimiento que se sigue en cada caso es distinto, como se verá a continuación.

A continuación se va a hacer una explicación de todas las técnicas de análisis univariante que se han aplicado.

2.1. Prueba de la T de Student

Esta prueba se utiliza para comprobar la igualdad de las medias de dos muestras, o si la media de una muestra es igual a una media teórica determinada.

En el caso de este estudio, se van a realizar dos tipos de pruebas distintos: una para datos apareados, y otra para muestras independientes. En el caso de la comparación entre tiempos se usará la prueba para datos apareados, ya que de cada una de las madres se tiene información de ambos tiempos, por lo que se tienen las medidas de una misma variable dos veces para cada madre (una medida a día 7 y otra a día 15). En cambio, en el caso de la comparación entre las dos dietas, se va a realizar la prueba para muestras independientes, ya que cada una de las madres ha seguido uno de los dos tipos de dietas, por lo que de cada madre solo se tendrá una medida.

Aun así, el proceso que se sigue para la realización de la prueba es el mismo. Primero se deben establecer las hipótesis a testear: la hipótesis nula, como se ha comentado anteriormente, es la que considera que no hay diferencia o cambio, por lo que afirma que no hay diferencias entre los grupos. Por otra parte, la hipótesis alternativa considera que el valor real de la media de un grupo es distinto (o superior o inferior, aunque en el caso de este estudio no se usarán) al valor real de la media del otro grupo.

El segundo paso es el del cálculo del estadístico. En el caso de las muestras independientes, el estadístico t se calcula de la siguiente forma:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \quad (1)$$

Donde \bar{X}_1 y \bar{X}_2 son las medias en cada grupo; σ_1^2 y σ_2^2 son las varianzas del primer y el segundo grupo, respectivamente; y N_1 y N_2 es el número de observaciones en cada grupo.

En cambio, en el caso de las muestras apareadas, para calcular el estadístico se debe, primero, calcular la diferencia entre las dos muestras (es decir, obtener una nueva variable que sea la diferencia entre la medida del primer grupo y la del siguiente). Una vez obtenida esta diferencia, la fórmula a seguir para el cálculo del estadístico es la siguiente:

$$t = \frac{\bar{X}_D}{\frac{S_D}{\sqrt{N}}} \quad (2)$$

Donde \bar{X}_D es la media de las diferencias; S_D es la desviación estándar de las diferencias; y N es el número de pares de observaciones.

Una vez se tiene el valor del estadístico, se puede proceder buscando dicho valor en la tabla de la T de Student, en función de los grados de libertad que se tengan. Dichos grados de libertad se calculan, en el caso de las muestras independientes, como $(N_1 - 1) + (N_2 - 1)$ y, en el caso de muestras apareadas, como $N - 1$.

Por lo tanto, basta con relacionar los grados de libertad con el nivel de significación que se esté considerando en la tabla de la T de Student, y comparar dicho valor con el del estadístico obtenido. Con esto encontraremos el valor crítico para rechazar la hipótesis nula, por lo que si el valor del estadístico t que se ha obtenido es mayor que dicho valor crítico, se rechazará la hipótesis nula.

Aun así, esta prueba se puede aplicar muy fácilmente con el software R, usando la función `t.test`. En ella, además, se puede especificar si se está realizando una prueba para muestras independientes o apareadas, mediante el parámetro `paired =` , que será TRUE si la prueba es para datos apareados y FALSE en caso contrario.

2.2. Modelos de Tobit

Los modelos de regresión Tobit son modelos lineales para datos censurados. Por lo tanto, antes de entrar en el detalle de cómo funciona un modelo Tobit, es conveniente definir qué son los datos censurados.

Se considera que los datos están censurados cuando existe un determinado límite en la variable respuesta a partir del cual a todas las observaciones se les asigna un mismo valor. Hay distintos motivos por los que podemos encontrar datos censurados; uno de ellos es que el instrumento de medida tiene un límite de detección por debajo de lo cual todo valor se considera cero.

En el caso de este estudio, por definición las variables correspondientes al grupo de las citoquinas tienen una concentración muy baja en la leche materna, como ya se ha visto en apartados anteriores; éste es el motivo de que la mayor parte de las medidas que se toman sobre variables de este grupo sean iguales a cero.

Cuando eso pasa, debido al gran número de valores cero que se tienen, no es conveniente realizar una prueba como la T de Student, ya que los valores iguales a cero pueden provocar un sesgo muy grande, ya que pueden hacer que, a pesar de que la variable que nos interesa estudiar tenga una distribución determinada, los valores que realmente se observen en la muestra no sean representativos. Es por eso por lo que se ha propuesto la alternativa que se presenta en esta sección, que son los modelos de Tobit.

El objetivo del modelo es describir la relación entre una variable dependiente no negativa y una variable independiente; en el caso de este estudio, por lo tanto, queremos describir la relación entre cada una de las variables del grupo de las citoquinas y los distintos tiempos o las distintas dietas, según el objetivo.

Lo que se hace en estos modelos es definir una nueva distribución para la variable censurada a partir de una variable aleatoria original subyacente, que se denominará como variable latente, y que aquí se denotará como y_i^* .

Esta variable latente depende linealmente de la variable independiente que se considere para el modelo (x_i) a través de un vector que determina la relación entre ellas, y que denotaremos β . Además, también se tiene en cuenta un término de error u_i para captar las influencias aleatorias de la relación.

Por lo tanto, la variable latente se definirá como:

$$y_i^* = \beta x_i + u_i \quad (3)$$

Es importante tener en cuenta que el coeficiente β no se interpreta como el efecto de x_i sobre y_i , si no que se interpreta como la combinación del cambio en y_i de aquellos por encima del límite, ponderados por la probabilidad de estar por encima del límite, y el cambio en la probabilidad de estar por encima del límite, ponderado por el valor esperado de y_i si es superior ⁴.

⁴https://es.wikipedia.org/wiki/Modelo_Tobit

A partir de esta variable latente, se obtendrá la nueva distribución de la variable censurada, como:

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \quad (4)$$

De esta forma, la nueva variable tomará el valor de la variable latente en caso de que ésta sea superior a cero, y cero en caso contrario.

Este modelo se puede aplicar en R usando la función `tobit`. En este caso, no hay un parámetro concreto para definir si la muestra es apareada o no lo es; por lo tanto, en el caso de las muestras apareadas, lo que se hace es calcular una nueva variable con la diferencia de las medidas en los distintos grupos (los dos días, en el caso de este estudio) y hacer la inferencia con una regresión que tan solo tiene en cuenta el término independiente. Por lo tanto, la notación a usar en el caso de muestras apareadas sería la siguiente: `tobit(dif~~1,left=0,dist='gaussian')`; y la que se usaría en el caso de muestras independientes sería `tobit(variable dependiente~~variable grupo,left=0,dist='gaussian')`.

En la formulación que se ha especificado, el parámetro *left* corresponde al límite inferior de la censura de la variable dependiente, y *dist* a la distribución que se asume para la variable dependiente, que en este caso se ha determinado como la gaussiana.

2.3. Prueba de la Chi Cuadrado

Para aquellas variables categóricas y/o binarias del estudio, como las infecciones en distintos tiempos, se ha usado una prueba de independencia basado en la Chi Cuadrado, que consiste en la comparación de proporciones.

En estas pruebas, la hipótesis nula que se plantea es que no existen diferencias entre las proporciones de las poblaciones. Por lo tanto, sigue el mismo esquema que el visto con las dos pruebas anteriores: si una variable es significativa, querrá decir que hay diferencias entre las proporciones de los distintos grupos.

El estadístico que permite hacer dicho contraste es se basa en la Chi cuadrado, y se basa en los valores observados y los esperados. Los esperados se calculan multiplicando las frecuencias marginales y dividiendo el producto por n. Estas E_{ij} son estimaciones de las frecuencias absolutas que cabría esperar en cada casilla bajo el supuesto de que la proporción de éxitos es la obtenida a partir del total de observaciones muestrales sin considerar diferencias entre los dos grupos.

El cálculo del estadístico se hace mediante la siguiente fórmula:

$$\chi^2 = \sum_{j=1}^2 \sum_{i=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (5)$$

Donde O_{ij} se refiere a los valores observados y E_{ij} a los esperados. Además, i recorre los individuos del primer grupo y j los del segundo, de forma que I es el número total de individuos en el primer grupo y J el número de individuos del segundo.

Dicho estadístico presenta una distribución Chi cuadrado con $(I - 1) * (J - 1)$ grados de libertad. El proceso para determinar el rechazo o aceptación de la hipótesis nula es el mismo que en el caso de la T de Student, aunque en este caso haciendo uso de la tabla de la Chi Cuadrado: se relacionan los grados de libertad con el nivel de significación que se esté considerando en la tabla, y se compara dicho valor con el del estadístico obtenido, encontrando así el valor crítico para rechazar la hipótesis nula, de forma que si el valor del estadístico es mayor que el valor crítico, se rechazará la hipótesis nula.

De nuevo, esta prueba se puede hacer mediante R usando la función `prop.test`. En este caso, la función necesita dos parámetros: la tabla de contingencia entre las dos variables (la variable dependiente y la variable con los grupos) y la tabla de frecuencias para la variable dependiente.

Otra opción para la realización de la prueba desde R es mediante la función `chisq.test`; en este caso, tan solo hace falta especificar la tabla de contingencia entre las dos variables que se quieren comparar.

2.4. Prueba de Fisher

Igual que la prueba de la Chi Cuadrado, la prueba de Fisher se usa cuando se quiere estudiar si existe asociación entre dos variables cualitativas (si las proporciones de una variable son diferentes dependiendo del valor que tenga la otra variable). En este caso, es un test exacto, y es más preciso que la prueba de la Chi Cuadrado u otros tests de independencia cuando el número de eventos esperado por cada grupo es pequeño.

Para poder realizar esta prueba, se deben cumplir las condiciones de aleatoriedad (el muestreo debe haber sido aleatorio), del tamaño de la muestra (que debe ser menor al 10% del total de la población) y además cada observación debe contribuir únicamente a uno de los niveles.

En este caso, el p-valor se calcula a partir de la tabla de contingencia entre ambas variables categóricas. Concretamente, se calculará como:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!} \quad (6)$$

Donde el signo de exclamación (!) se refiere al operador factorial, a se corresponde con el recuadro superior izquierdo de la tabla de contingencia (por lo tanto, el número de individuos que son del primer grupo de la primera variable y del primer grupo de la segunda), b el recuadro superior derecho (por lo tanto, el número de individuos que son del primer grupo de la primera variable y del segundo grupo de la segunda), c el inferior izquierdo (por lo tanto, el número de individuos que son del segundo grupo de la primera variable y del primer grupo de la segunda) y d el inferior derecho (por lo tanto, el número de individuos que son del segundo grupo de la primera variable y del segundo grupo de la segunda).

En R se puede realizar esta prueba mediante la función `fisher.test`, y solo necesita como parámetro la tabla de contingencia entre las dos variables categóricas.

2.5. Suma de rangos de Wilcoxon

Esta prueba no paramétrica también se puede llamar como la prueba de la U de Mann-Whitney. Como ya se ha comentado, es útil cuando se tienen dos muestras independientes. Por lo tanto, son muestras que pueden no tener el mismo tamaño. La prueba se usa para comprobar la heterogeneidad de dos muestras, es decir, si las dos muestras proceden de poblaciones equidistantes (si se pueden considerar iguales o no).

La idea principal de la prueba es que, si proceden de la misma población, al juntar todas las observaciones y ordenarlas de menor a mayor, cabría esperar que las observaciones de ambas muestras estuviesen intercaladas aleatoriamente, en lugar de agruparse de modo que las observaciones de una de las pruebas quedasen por encima de las de la otra.

Para realizar esta prueba, primero se deben juntar todas las observaciones, ordenarlas y asignarles un número de posición (des del 1 hasta el número de observaciones que se tengan). En caso de haber valores repetidos, se les asigna como valor de posición la media de las posiciones que ocupan. Seguidamente se calcula el rango para cada uno de los grupos, como el sumatorio de las posiciones que pertenezcan a cada grupo.

A continuación se calculan los estadísticos U , en base al número de observaciones de cada muestra y del rango. En concreto, se calculará un estadístico para cada muestra, de la siguiente manera:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad (7)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad (8)$$

Donde n_1 y n_2 son el número de observaciones de la primera y la segunda muestra, respectivamente, y R_1 y R_2 el sumatorio de los rangos de la primera y la segunda muestra, respectivamente.

Una vez se ha obtenido el estadístico para cada uno de los grupos, se procede al cálculo del estadístico U final, como el mínimo entre los dos estadísticos calculados anteriormente.

El proceso para determinar si se debe rechazar o no la hipótesis nula es similar al que se ha visto en las pruebas paramétricas. Se debe mirar el valor de la tabla U de Mann-Whitney, y si el estadístico U es menor que el valor correspondiente en la tabla, la diferencia entre las dos muestras es significativa (y, por lo tanto, hay diferencias entre los dos grupos).

En R se puede realizar este test mediante la función `wilcox.test`, especificando las dos muestras y si se quieren tener en cuenta las dos colas (de forma que se mirará si las dos muestras son diferentes) o solo una de ellas (comprobando si una es mayor o menor que la otra), cosa que se puede especificar mediante el parámetro *alternative*, que en este caso será igual a *two-sided* ya que queremos comprobar la diferencia de muestras.

2.6. Prueba de rangos y signos de Wilcoxon

Como ya se ha comentado, esta prueba es el equivalente a la prueba T de Student para datos apareados, aunque en lugar de trabajar con las medias trabaja con las medianas.

Para llevarla a cabo, igual que pasaba en la prueba de la T de Student primero se debe calcular la diferencia en valor absoluto de cada par de observaciones. Seguidamente, se ordenan estas diferencias y a cada una se le asigna el valor de su posición, des de 1 hasta el número de pares de observaciones que se tengan. Igual que en la prueba de suma de rangos, en caso de haber empates se les asigna la media de las posiciones que ocupan.

Además, se crea otra variable con el signo de la diferencia: si es más grande el valor de la primera muestra, tomará valor 1, y en caso contrario, -1.

A continuación, se calcula la suma de rangos con signo positivo (1) y, por otro lado, la suma de rangos con signo negativo (-1). Con eso, se calcula el valor del estadístico W , como el mínimo entre el sumatorio de los rangos con signo positivo y la suma de los rangos con signo negativo.

En este caso, se puede determinar el rechazo o aceptación de la hipótesis nula de dos formas distintas: la primera, que se usa para tamaños de muestra pequeños (menores a 25 observaciones) se compara el valor del estadístico con el valor de la tabla de Wilcoxon, y si el estadístico cae dentro del intervalo correspondiente en la tabla para ese número de observaciones, la diferencia no es significativa; y la segunda opción, que es la que usa R , es asumir que el estadístico se distribuye de forma aproximadamente Normal, de forma que se calcula un nuevo estadístico Z y se compara éste con el valor de la tabla Normal para el nivel de significación que se ha escogido, rechazando la hipótesis nula si el estadístico calculado es menor que dicho valor.

En R se puede realizar esta prueba mediante la misma función que en el caso de la suma de rangos (`wilcox.test`), aunque en este caso se debe especificar el parámetro `paired = TRUE`, y no se contempla el parámetro referente a la alternativa.

3. Análisis multivariante

Para realizar un análisis multivariante de los datos, que sirve para la observación y el análisis simultáneos de más de una variable respuesta, existen distintas técnicas. En el caso de este estudio, se ha decidido utilizar las técnicas del Análisis de Factores Múltiples y el Random Forest.

3.1. Análisis de Factores Múltiples

El Análisis de Factores Múltiples es una técnica estadística útil para analizar de forma simultánea diversas variables. Por lo tanto, es un método de análisis de datos multivariantes, que sirve para resumir y visualizar una tabla de datos compleja en la que los individuos son descritos por varios conjuntos de variables (cuantitativas y / o cualitativas) estructuradas en grupos, teniendo en cuenta la contribución de todos los grupos activos de variables para definir la distancia entre los individuos.

El número de variables en cada grupo puede diferir y la naturaleza de las variables (cualitativa o cuantitativa) puede variar de un grupo a otro, pero las variables deben ser de la misma naturaleza en un grupo determinado.

Este análisis global, donde se consideran simultáneamente múltiples conjuntos de variables, requiere equilibrar las influencias de cada conjunto de variables.

Este método de análisis puede ser considerado una generalización de las técnicas de Análisis de Componentes Principales (PCA) y de Análisis de Correspondencias Múltiples (MCA). De hecho, su núcleo de MFA se basa en un análisis factorial (PCA en el caso de variables cuantitativas, MCA en el caso de variables cualitativas) en el que las variables se ponderan. Estos pesos son idénticos para las variables del mismo grupo (y varían de un grupo a otro).

Técnicamente, MFA asigna a cada variable del grupo j , un peso igual al inverso del primer valor propio del análisis (PCA o MCA según el tipo de variable) del grupo j . El objetivo del análisis es obtener una imagen integrada de las observaciones y de las relaciones entre los grupos de variables.

Una de las aplicaciones del MFA la tenemos con los datos longitudinales, donde los individuos han sido medidos en diversas ocasiones a lo largo del tiempo. En este contexto existen dos posibles enfoques:

- Crear los grupos con las variables que se han medido en un mismo tiempo. De esta forma cada grupo representa un instante temporal.
- Crear un grupo para cada variable con todas las medidas de esa variable a los diferentes tiempos.
- Si tratamos con dos tiempos se puede calcular el incremento y representar los diferentes grupos de variables respecto al incremento.

El MFA puede ser complementado por un grupo de variables suplementarias que no intervienen en la definición de los ejes pero pueden representarse en el espacio generado por ellos para visualizar una posible relación entre las variables suplementarias y las dimensiones del MFA representadas por los ejes.

En R, se puede realizar este análisis mediante la función `MFA`. Con ello, obtendremos un conjunto de gráficos que nos servirán para llegar a distintas conclusiones sobre el análisis.

Un primer paso que debe seguirse es el de escoger el número de dimensiones con las que se quiere trabajar. Cada dimensión explicará un determinado porcentaje de variabilidad de los datos, y queda a elección del autor seleccionar el criterio a seguir para determinar el número de dimensiones con las que trabajar. En este estudio, como se verá, las dimensiones no explicaban un porcentaje demasiado alto de variabilidad, por lo que nos hemos conformado con que se explicara como mínimo el 50 % de dicha variabilidad de los datos. Para saber el porcentaje de variabilidad que cada dimensión explica, se puede aplicar la función `fviz_screeplot` sobre el objeto MFA creado. Dicha función devuelve un histograma que representa los porcentajes para cada dimensión.

Por otro lado, mediante la función `fviz_mfa_ind` se consiguen los gráficos de los individuos. Esta función tiene un parámetro (`axes`) para determinar qué par de dimensiones se quieren graficar (ya que al ser un gráfico en un plano solo es posible graficar dos dimensiones a la vez). Es por eso por lo que se grafican la primera y segunda dimensiones en un gráfico, la tercera y la cuarta en otro, y así consecutivamente hasta haber graficado todas las dimensiones escogidas. El gráfico que se obtiene mediante esta función representa cada uno de los individuos del estudio sobre las dimensiones; por lo tanto, se puede observar dónde se posicionaría cada uno de ellos en el plano. Esto es útil para observar patrones de comportamiento entre individuos.

Además, se pueden graficar los grupos sobre las dimensiones (es decir, observar dónde se posiciona cada uno de los grupos con respecto a las dimensiones). Esto nos ayudará a saber con qué dimensión está más relacionado cada uno de los grupos, o dicho de otra forma, cuál es el grupo que mejor representa cada dimensión, ya que cuanto más cerca del eje de la dimensión esté un grupo, más relación tiene con ésta.

Otro gráfico que es útil para llegar a conclusiones relevantes es el gráfico de individuos parciales. En él se representa cada uno de los grupos en su posición sobre las dimensiones como en el caso de los gráficos de individuos, pero en este caso se añade también la posición en la que estaría cada individuo en el caso de tener en cuenta solo uno de los grupos. De esta forma, de cada individuo salen tantas flechas como grupos haya, y en el extremo de dicha flecha aparece un punto que sería la posición del individuo sobre las dimensiones en caso de tener en cuenta solo aquél grupo.

Es importante recalcar, además, que no todas las variables tienen el mismo peso sobre las dimensiones, por lo que es relevante observar también la contribución de cada una de ellas sobre las dimensiones, así como la correlación entre cada variable y cada una de las dimensiones.

Adicionalmente, en el gráfico de los individuos se pueden colorear en función de sus características - por ejemplo, aplicar un color distinto en función de la dieta que cada uno siga, o en función de si ha sufrido o no una infección. Con esto se puede ver si los individuos con las mismas características se agrupan en el plano de las dimensiones y, por lo tanto, se comportan de forma similar, o no.

De esta forma, con el conjunto de información que se puede observar sobre un objeto MFA, es posible observar comportamientos distintos entre los grupos e incluso entre individuos. Por lo tanto, eso nos es útil para comprobar si hay una diferencia global entre los grupos, y no tan solo variable a variable como en el caso de las técnicas anteriores.

3.2. Random Forest

El método de Random Forest es un algoritmo de aprendizaje supervisado de predicción.

Decimos que es un método de aprendizaje supervisado porque los resultados que se obtienen son conocidos previamente; es decir, en el conjunto de datos se tiene una característica objetivo para cada uno de los registros, que será la que se quiere reproducir con el modelo.

Por otra parte, es un método de predicción porque consiste en la elaboración de un modelo predictivo: es útil para la clasificación, y también lo es para la regresión, permitiendo predecir tanto una variable categórica (en el caso de la clasificación) como una variable numérica (en el caso de la regresión).

Un Random Forest consiste en la elaboración de un conjunto de árboles de decisión, combinándolos con *bagging*. El *bagging* es una de las formas que existen para construir un conjunto de modelos de *machine learning*; cada uno de ellos produce una predicción diferente, y lo que hace el *bagging* es combinarlos para obtener una única predicción. En este método se consigue que los errores se compensen entre sí entrenando cada modelo con subconjuntos del conjunto de entrenamiento, eligiéndolos aleatoriamente.

Por lo tanto, cuando se realiza un Random Forest cada árbol que se crea está basado en un conjunto de datos distinto, por lo que ningún árbol ve todos los datos de entrenamiento. De esta forma, al combinar los resultados de los distintos árboles, unos errores se compensan con los otros y la predicción generaliza mejor. Dicho de otra forma, con el Random Forest conseguimos mejorar el error de generalización con respecto al que obtendríamos realizando un árbol de decisión.

Las predicciones de cada uno de los "sub-árboles" se combinan de forma distinta en los modelos de clasificación que en los de regresión. En los primeros, se suele usar el método del voto suave, que da más importancia a los resultados en los que los árboles estén muy seguros. En cambio, en los problemas de regresión se suelen combinar tomando su media aritmética.

En resumen, mediante el Random Forest seremos capaces de obtener un modelo predictivo, que en este caso servirá para poder determinar qué factores son los que provocan que haya dos grupos de madres bien diferenciados, como se verá más adelante.

En R se puede realizar esta técnica mediante la función `randomForest`, que pertenece a un paquete con el mismo nombre. Tan solo tenemos que especificar el modelo - la variable independiente (la que queremos predecir) y las variables independientes - y el número de árboles a realizar.

El Random Forest, pero, no lidia bien con los valores faltantes. Existen varias formas de solucionar este problema, pero en el caso de este estudio, se ha considerado el uso de la función `rfimpute`, que imputa automáticamente los valores faltantes de una base de datos. En concreto, imputa las variables numéricas a la media ponderada de las observaciones que no tienen valores faltantes, y las variables categóricas a la categoría con una proximidad media mayor, iterando el proceso varias veces hasta conseguir una imputación completa de los datos.

Además, y dado el bajo número de observaciones que se tienen, se ha decidido hacer una validación de los resultados mediante el método de *Leave-one-out*. En concreto este método consiste en entrenar todos los datos excepto en un punto, y hacer una predicción para este punto. Esto nos servirá para comprobar que la bondad de las predicciones obtenida mediante el Random Forest se mantendría si cogiésemos otra combinación de datos de entrenamiento y prueba.

IV. Resultados

En este apartado, se van a presentar los resultados más relevantes que se han obtenido en los dos objetivos que se han introducido, a los que se ha llegado mediante las metodologías explicadas en el apartado anterior.

Objetivo 1. Comparación entre los dos tiempos

Como se ha comentado anteriormente, en este objetivo del estudio se quiere observar si hay diferencias en la composición de la leche a día 7 y a día 15.

Para ello, se han seleccionado las variables referentes a los valores inmunológicos, que son los que determinan como está compuesta la leche, y se han realizado diversas pruebas para poder comprobarlo.

Análisis descriptivo inicial y transformación de los datos

En un primer paso, se ha hecho un análisis descriptivo inicial de los datos, en el que se ha podido observar que la mayoría de citoquinas tienen un porcentaje de ceros muy elevado, superior al 80 % en muchas de las variables correspondientes a ese grupo; aun así, eso no sorprende, ya que como se ha comentado son valores que no se presentan en una gran concentración en la leche materna. En cambio, las inmunoglobulinas y las adipocinas presentan un porcentaje de cero muy pequeño.

Además, en este análisis descriptivo inicial se ha observado el comportamiento de cada una de las variables mediante el cálculo de estadísticos (como la media, la desviación estándar o la mediana) y mediante una gráfica de boxplots múltiples. Con eso se ha podido observar que las variables, en general, no siguen una distribución normal, ya que no están centradas. Es por eso por lo que se ha decidido transformar los datos mediante una transformación logarítmica. En concreto, cada una de las variables del estudio se ha transformado siguiendo la fórmula que se presenta a continuación:

$$variable = \log(variable + 1) \quad (9)$$

Haciendo esta transformación, y aunque hay algunas variables que aún presentan cajas asimétricas, en general conseguimos bastante simetría, y se ha conseguido una distribución de las variables mucho más centrada, de forma que podemos aceptar que las variables siguen una distribución Normal, que es lo que nos conviene para poder hacer pruebas paramétricas. Es relevante comentar que, en las citoquinas, hay bastantes cajas que no tienen ni siquiera amplitud; eso se debe al gran número de valores ceros que tienen las variables correspondientes a dichos gráficos.

Comparación entre los dos tiempos

Una vez se ha realizado la transformación de los datos se puede proceder a la comparación entre el día 7 y el día 15 de todas las variables. Dicha comparación se ha hecho tanto mediante un método gráfico - graficar boxplots de cada variable en ambos tiempos -, como numéricos - una prueba paramétrica (prueba de la T de Student o modelo Tobit) y una prueba no paramétrica (prueba de Wilcoxon) -.

En concreto, para las inmunoglobulinas, las adipocinas y las variables calculadas se ha realizado la prueba T de Student para datos apareados, y una prueba no paramétrica, en este caso la prueba de Wilcoxon para datos apareados. En cambio, para las citoquinas, aunque se ha realizado la misma prueba no paramétrica, se ha decidido sustituir a prueba T por un modelo tobit, debido a la gran cantidad de ceros que tienen las variables de este grupo.

Para las inmunoglobulinas, y a pesar de que hemos visto que podemos considerar que la distribución de los datos es normal (de forma que la conclusión de la que nos fiáramos es la de la prueba T), en este caso vemos

que la prueba paramétrica y la no paramétrica coinciden en sus conclusiones, ya que en ambas salen las mismas variables como significativas: IgM, IgG2, IgE e IgA. Eso quiere decir que, para estas variables, no se puede considerar que la diferencia entre los dos tiempos sea igual a cero y, por lo tanto, podemos concluir que hay diferencias entre las medidas del día 7 y las del día 15. En cambio, en las variables IgG1, IgG3, IgG4 e IgG no podemos hacer tal afirmación, si no que debemos aceptar que no hay diferencias entre ambos tiempos de estudio.

Por otro lado, en las citoquinas sí que se observan diferencias entre las conclusiones del modelo Tobit y la prueba de Wilcoxon; el primero concluye que hay ocho variables con un p-valor significativo, lo que quiere decir que se puede considerar que hay diferencias entre los dos tiempos. En cambio, mediante la prueba no paramétrica tan solo hay cuatro variables significativas (una de las cuales no lo es mediante el modelo Tobit). Aun así, y ya que el modelo Tobit es capaz de comprender la gran cantidad de ceros de los datos, son más fiables para el estudio los resultados de este modelo, que concluye que en las variables IL5, IL17, IL18, IL10, IL6, IL9, IL1b y Ifn.g hay diferencias entre la medida del día 7 y la medida del día 15.

En cuanto a las adipoquinas, ambas variables tienen p-valores muy grandes en las dos pruebas y, por lo tanto, se confirma que no hay diferencias entre el día 7 y el día 14 en este grupo de variables.

Finalmente, en el caso de las variables calculadas las dos pruebas difieren en dos de las variables: Th17, donde la prueba no paramétrica determina que hay diferencias entre tiempos y la paramétrica no, y Acquired Immunity, donde pasa al revés. Viendo los boxplots comparativos, en ambos casos podríamos decir que hay diferencias entre el gráfico del día 7 y el gráfico del día 15; además, en los casos no significativos el p-valor queda bastante al límite del nivel de significación, siendo 0.063 y 0.066 respectivamente. Así pues, podríamos determinar que en ambas variables hay diferencias entre los tiempos. Las variables en las que dicha diferencia sí está clara, por tener un p-valor significativo en ambas pruebas, son Cito_Th2, Innate Immunity y Anti_infl.

A continuación se presenta una tabla que muestra aquellas variables que sí que han resultado significativas y en las que, por lo tanto, se observan diferencias entre los dos tiempos.

	Variable	IC 95 % / Q1-Q3	p-valor (paramétrico)	p-valor (no paramétrico)
Inmunoglobulina	IgM	(0.011, 0.805)	0.044	0.032
	IgG2	(0.066, 0.558)	0.014	0.005
	IgE	(0, 0.002)	0.006	0.002
	IgA	(0.218, 0.508)	3.93×10^{-6}	1.78×10^{-11}
Citoquinas	IL5	0	0.048	0.379
	IL17	0	0.019	0.163
	IL18	(-0.465, 1.307)	0.009	0.030
	IL10	(-0.142, 0.04)	0.006	0.412
	IL6	0(-2.491, 0.253)	0.006	0.009
	IL9	0	0.005	0.239
	IL1b	(-0.855, 0)	0.004	0.015
	Ifn.g	0	0.003	0.216
Variables calculadas	Acquired Immunity	(0.029, 0.563)	0.030	0.066
	Th2 (Citoquina)	(0.039, 0.555)	0.025	0.026
	Innate Immunity	(0.209, 1.158)	0.005	0.004
	Anti Infl	(0.239, 1.293)	0.005	0.006

Tabla 1: Variables con diferencias entre el día 7 y el día 15 ⁵

Análisis de Factores Múltiples

Para ver si hay comportamientos distintos entre los dos tiempos de estudio, se han realizado dos análisis distintos: para los dos tiempos (es decir, el día 7 y el día 15); y un segundo teniendo en cuenta incrementos

⁵Elaboración propia.

en el tiempo.

Análisis de Factores Múltiples para dos tiempos

Para la realización de este análisis, se ha elaborado un análisis de factores múltiples teniendo en cuenta cada tiempo como un grupo. De esta forma, se ha realizado un estudio con el que se podrá ver la posición de cada variable respecto a cada tiempo.

El primer paso para el análisis de factores múltiples es el de decidir el número de dimensiones con las que se va a trabajar. En este caso, la primera dimensión consigue explicar el 16.9% de la variabilidad de los datos. Por su parte, la segunda dimensión explicaría el 12.9% de la variabilidad, mientras que la tercera y cuarta dimensiones explicarían el 11% y el 7.4%, respectivamente. A partir de esta dimensión, el porcentaje de variabilidad explicado se mantiene más constante en el resto de las dimensiones. Es por eso por lo que, en este caso, hemos decidido trabajar con cuatro dimensiones, que conjuntamente, consiguen explicar el 48% de la variabilidad.

Una vez escogido el número de dimensiones, se puede proceder a la representación de los individuos, que será el que nos ayudará a llegar a las conclusiones que correspondan. Primeramente, se graficaron los individuos sobre cada una de las dimensiones (al no poder graficarlos sobre las cuatro dimensiones a la vez, se dividieron en dos: uno para la primera y segunda dimensiones, y otro para la tercera y la cuarta).

En el gráfico de la primera y segunda dimensiones, se observan claramente dos clústeres distintos. En cambio, en la tercera y cuarta dimensiones se observa tan solo un grupo. En ambos casos hay individuos alejados de cualquiera de las nubes de puntos; dichos individuos, como el 75, el 30 o el 22 en las dos primeras dimensiones o el 79 en la tercera y cuarta, son aquellos que tienen comportamientos más extremos, y se pueden identificar como outliers; si dichos individuos no estuviesen en el estudio, eso afectaría a la variabilidad de los otros individuos, por lo que los gráficos que obtendríamos serían distintos.

Dicho gráfico en el que se pueden observar los dos grupos bien identificados es el que se muestra a continuación:

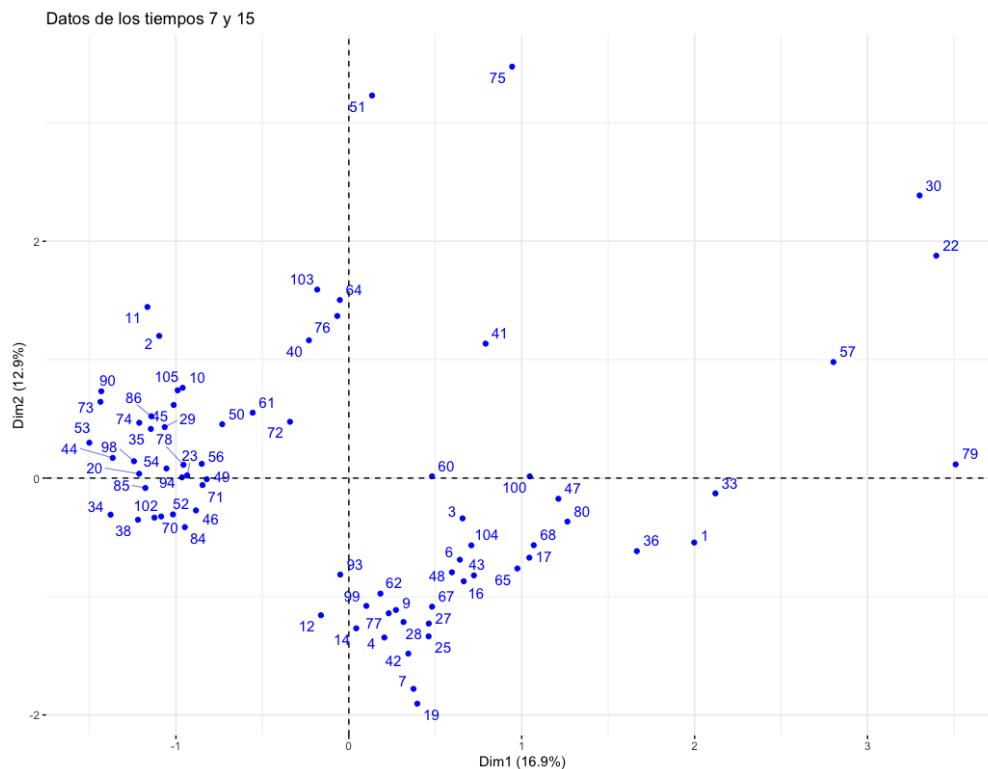


Figura 2: Gráfico de los individuos en la primera y segunda dimensiones (MFA de dos tiempos) ⁶

Es interesante comprobar si los clústeres que se han detectado en el gráfico de la primera y segunda dimensiones se mantienen cuando separamos las variables por tiempo o si, por lo contrario, se necesitan ambos tiempos para detectarlos. Para ello, se han separado las variables pertenecientes al día 7 de aquellas del día 15, y se han vuelto a graficar los individuos. Los gráficos pertenecientes a dicho análisis son los siguientes:

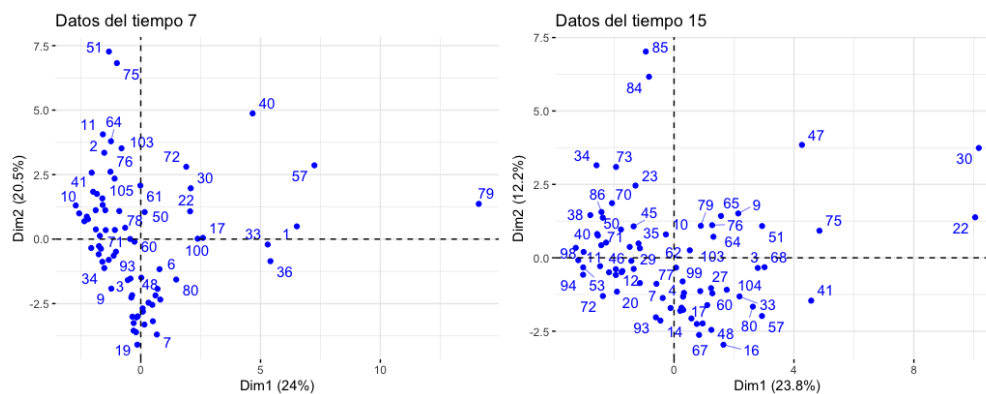


Figura 3: Gráfico de los individuos en la primera y segunda dimensiones en el día 7 (MFA de dos tiempos)
7

Con los gráficos obtenidos, se puede observar que no se mantienen los clústeres que se habían detectado. Por lo tanto, es necesario tener ambos tiempos para poder discriminar, lo que quiere decir que el comportamiento de los individuos a día 7 y a día 15 es distinto.

En cuanto a la representación de los grupos sobre cada eje, los tiempos quedan representados de forma muy distinta según sobre qué eje se grafiquen. La contribución del tiempo 15 es más grande que la del tiempo 7 en la primera y cuarta dimensiones (habiendo mucha más distancia respecto a la contribución del tiempo 7 en la cuarta dimensión), mientras que en la segunda y tercera dimensión pasa al revés (el tiempo 7 tiene más contribución sobre estas dimensiones). En caso de que mirásemos la contribución de cada tiempo sobre la primera y segunda dimensión conjuntamente, veríamos que el tiempo 7 es el que más contribuye; por otro lado, si hiciésemos lo mismo con la tercera y cuarta dimensiones, sería el tiempo 15 el que más contribuiría. Aun así, es relevante comentar que, en las dos últimas dimensiones, la contribución de los tiempos es más parecida, mientras que en las primeras el tiempo 7 contribuye bastante más.

Además, se ha hecho la representación parcial de los individuos, con la que se puede ver dónde se ubica cada individuo y también dónde se posicionaría en caso de solo tener en cuenta uno de los dos tiempos. Por lo tanto, estos gráficos indican el movimiento entre tiempos. Gracias a ellos se puede ver que realmente hay una diferencia significativa en el comportamiento de los individuos en función de los tiempos, ya que en caso de tener tan solo la medida de uno de los dos tiempos, la posición de muchos de los individuos sería muy distante de la que toman actualmente o de la que tendrían en caso de tener en cuenta tan solo el otro tiempo. Esto se puede observar, sobre todo, en el gráfico de la primera y segunda dimensiones, en el gráfico que se muestra a continuación:

⁶Elaboración propia.

⁷Elaboración propia.

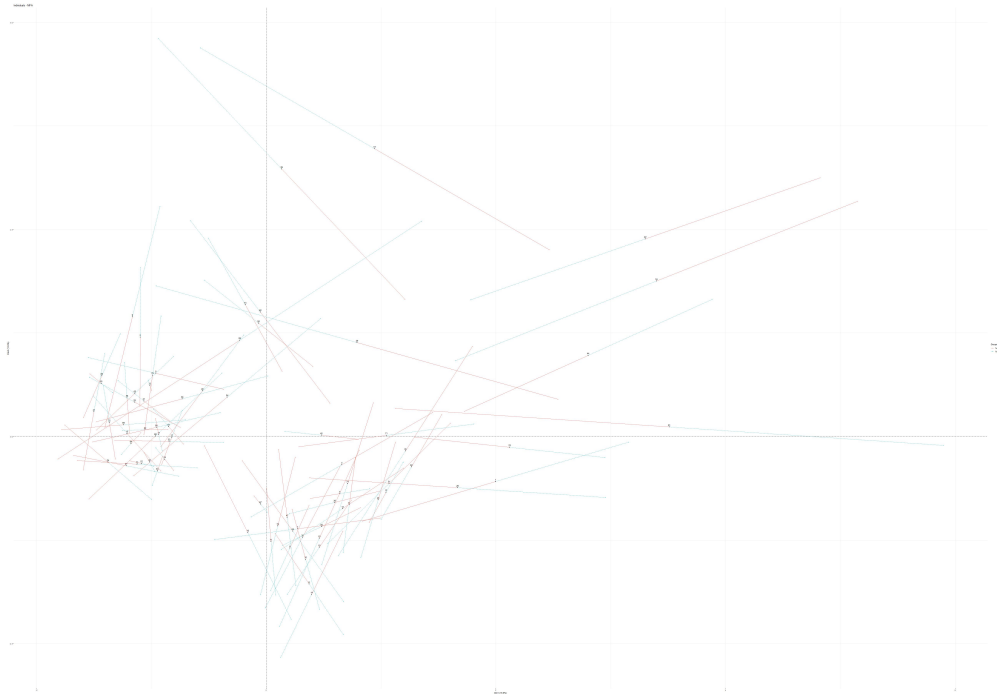


Figura 4: Representación parcial de los individuos en la primera y segunda dimensiones (MFA de dos tiempos) ⁸

Es relevante comentar que, en dicho gráfico, se puede observar que la mayoría de madres de la nube de puntos inferior han hecho una evolución creciente (es decir, en el día 7 tienen menos concentración y en el 15, más) y, en cambio, pasa lo contrario con las madres de la nube superior, que tienen un movimiento decreciente. Así pues, los dos grupos de madres que se visualizan en el gráfico tienen un comportamiento distinto entre ellas en el tiempo.

Finalmente, es importante tener en cuenta que no todas las variables han contribuido de igual forma en cada una de las dimensiones que se han estudiado. En concreto:

- La mayoría de las variables que más contribuyen en la primera dimensión son inmunoglobulinas en el día 15.
- Todas las variables más contribuyentes en la segunda dimensión son inmunoglobulinas en el día 7.
- Todas las variables más contribuyentes en la tercera dimensión son citoquinas en el día 7.
- Todas las variables más contribuyentes en la cuarta dimensión son citoquinas, en este caso en el día 15.

Análisis de factores múltiples para el incremento entre tiempos

En este caso, se ha hecho un análisis de factores múltiples a partir del incremento entre los dos tiempos. Para ello, se ha calculado el valor absoluto de la resta de la medida entre la variable en el tiempo 7 y la medida de la misma variable en el tiempo 15, y se ha realizado el análisis tomando como grupos los tres tipos de variables de los que disponemos – adipoquinas, inmunoglobulinas y citoquinas.

De nuevo, se ha decidido trabajar con cuatro dimensiones, que representan el 50 % de la variabilidad de los datos.

⁸Elaboración propia.

En el caso de este nuevo análisis, al analizar la representación los individuos no se ve ninguna agrupación como sí pasaba en el análisis de los dos tiempos. De hecho, de estos gráficos de individuos podemos sacar pocas conclusiones; tan solo sería relevante comentar que hay individuos outliers, como podría ser la madre 79, que queda distante de la nube de puntos tanto en la primera y segunda dimensiones como en la tercera y cuarta.

En cuanto a los grupos, de nuevo quedan representados sobre los ejes de forma distinta en función de las dimensiones sobre las que estemos trabajando. Las adipoquinas y las citoquinas contribuyen más sobre la primera dimensión, en este orden. Por su parte, en la segunda y tercera dimensiones las variables más contribuyentes son las inmunoglobulinas y las adipoquinas, casi en el mismo porcentaje. Por último, en la cuarta dimensión las citoquinas son las variables más contribuyentes. Si tenemos en cuenta la contribución en el total de la primera y segunda dimensiones, las adipoquinas son las que tienen un porcentaje de contribución mayor y, de hecho, es mucho mayor que el de los otros dos grupos. En cambio, si se analizan conjuntamente la tercera y cuarta dimensiones son las citoquinas las más contribuyentes, aunque con un porcentaje muy similar al del resto de grupos.

En cuanto a la representación parcial de los individuos, los gráficos que se obtienen en este caso no aportan mucha información, ya que se sobreponen los individuos y no se pueden sacar conclusiones visuales; es por eso por lo que no se van a incluir en este apartado. Sin embargo, se pueden visualizar en el Anexo.

Finalmente, observando la contribución de las distintas variables en cada dimensión se ha comprobado que en todas las dimensiones la variable que más contribuye forma parte del grupo de adipoquinas. Además, si se estudia la correlación de las variables con los ejes, como se ha hecho en el estudio anterior, se puede ver que las dos únicas variables con una correlación alta (\cos^2 superior al 50%) son la variable Adiponectin respecto a la dimensión 1, y la variable Leptin con respecto a la dimensión 2. Éstas son, de echo, las dos únicas variables que forman parte del grupo de las adipoquinas.

Así pues, la conclusión general que podemos sacar del análisis de factores múltiples que tiene en cuenta el incremento entre los tiempos es que las adipoquinas tienen mucha variabilidad dentro de dicho incremento, y absorben casi toda la variabilidad de las dimensiones, por lo que quizás están escondiendo el comportamiento de las otras variables.

Teniendo en cuenta que tan solo se dispone de dos variables del grupo de adipoquinas, se ha decidido realizar el mismo estudio eliminando el grupo de las adipoquinas, con el objetivo de ver si, haciendo eso, se puede sacar alguna conclusión más relevante con respecto a los otros grupos.

Para este análisis también se han tenido en cuenta cuatro dimensiones. Al hacer los gráficos de los individuos, se ha visualizado una sola nube de puntos en cada uno de ellos; por lo tanto, no se detectan comportamientos distintos en los individuos en los distintos grupos.

El grupo de las inmunoglobulinas contribuye más en explicar la primera y tercera dimensiones, mientras que el de las citoquinas explica un porcentaje superior en la segunda y cuarta dimensiones; por lo tanto, en este caso sí que se observa que hay grupos que dependen más en una u otra dimensión (recordemos que, incluyendo las adipoquinas, éstas eran el grupo más contribuyente en todas las dimensiones exceptuando la cuarta). Si se analiza la contribución global en la primera y segunda dimensiones, los porcentajes están muy igualados en ambos grupos, aunque es más contribuyente el de las citoquinas. En el caso de la tercera y cuarta dimensiones, hay más distancia entre ellos, siendo superior el de las inmunoglobulinas.

Viendo el gráfico parcial para los individuos, se observa que, en algunos de ellos, como en la madre 79 o la 40 en la primera y segunda dimensiones, el movimiento que se produce si solo se tiene en cuenta uno de los grupos es bastante grande; se detecta que en los individuos outliers este movimiento es mayor que en el caso de los individuos que están en el interior de la nube de puntos.

Por último, en cuanto a la contribución de las variables llama la atención que, a pesar que en la cuarta dimensión el grupo predominante son las citoquinas, la variable que consigue contribuir más en ella es una inmunoglobulina, en concreto la IgM, con una contribución muy superior a la del resto de variables y que casi duplica la de la segunda variable más importante.

Por lo tanto, con esta propuesta tampoco se ha conseguido visualizar comportamientos que nos indiquen que haya diferencias entre los grupos a partir de los incrementos en el tiempo. Los resultados del análisis, al no presentar conclusiones relevantes, no se incluyen en el cuerpo del trabajo, pero sí que se pueden visualizar en el Anexo.

Objetivo 2. Factores influyentes en la composición de la leche.

El segundo objetivo del estudio era el de ver qué factores podían influir en la composición de la leche materna. Para poder determinarlo se realizaron una serie de análisis que se verán a continuación.

Análisis descriptivo inicial y transformación de los datos

En este objetivo se incluyeron en el estudio variables que en el objetivo anterior no se habían tenido en cuenta referentes a las características de la madre y del bebé, tales como la dieta que siguió la madre o el peso del bebé (teniendo en cuenta su evolución en distintos días), entre otras.

Estas variables adicionales no son todas numéricas, como sí que pasa en el objetivo anterior. Contamos también con algunas variables categóricas, como la dieta que siguió la madre, el género del bebé, el tipo de parto, si fue el primer parto de la madre o no y la existencia de animales en el núcleo familiar, otras variables que se pueden considerar binarias, relacionadas con el hecho de haber tenido o no infecciones en distintos momentos del tiempo, y otras que denominaremos como contadores, ya que determinan el número de antibióticos o probióticos que la madre o el bebé tomaron.

En un primer análisis descriptivo sobre las nuevas variables numéricas, basado en el graficado de los boxplots múltiples y el cálculo de distintos estadísticos, se decidió que aquellas variables numéricas también debían transformarse para asegurar que seguían una distribución Normal, y se hizo en base al mismo criterio de transformación logarítmica que el presentado en el objetivo anterior.

Adicionalmente, se hizo un análisis descriptivo sobre aquellas variables de las que se tenía información en distintos momentos del tiempo (concretamente, el índice de masa corporal y el peso, de los que se tenía información a los 7 días, a los 15 días, al mes, a los seis meses, al año, a los 18 meses y a los dos años) y de las infecciones (de las que se tenía información a los 6, 12 y 24 meses).

Para ello, se graficó cada variable a lo largo del tiempo, tanto en nubes de puntos, añadiendo una línea de tendencia sobre la media, como mediante boxplots. Con ese estudio se pudo ver como, mientras el índice de masa corporal no parece tener una tendencia clara a lo largo del tiempo, el peso va subiendo claramente a lo largo de los meses. En cuanto a las infecciones, se pudo comprobar que la distribución de la variable en el medio año es muy similar a la que tiene en el año y, en cambio, a los dos años ya varia, pudiendo llegar a tener más infecciones, aunque la asimetría de los datos se mantiene en los tres tiempos.

¿Es la dieta un factor influyente?

En un primer momento, se consideró que aquella variable que más podían influir en la composición de la leche materna era el tipo de dieta. Concretamente, cada madre podía estar realizando un tipo de dieta, el tipo I o el tipo II, y se tenía la sospecha de que esto influía en la composición de la leche. Es por eso por lo que se decidió hacer un análisis similar al que se hizo en la comparación de los dos tiempos en el primer objetivo para ver si había diferencias entre los dos tipos de dieta.

Concretamente, se realizó un test de la T de Student y un test de Wilcoxon para las variables del grupo de inmunoglobulinas, adipoquinas y las nuevas variables numéricas añadidas en el estudio; por otra parte, para las citoquinas también se realizó un test de Wilcoxon, aunque en este caso se escogió como prueba paramétrica el modelo de Tobit, igual que en el otro objetivo. En el caso de las variables que se pueden considerar contadores se realizó un contraste de proporciones, y para las variables categóricas una prueba de la Chi Cuadrado y una prueba de Fisher.

Aun así, esta idea no dió los frutos que imaginábamos, ya que no se encontraron diferencias en ninguna de las variables referentes a las inmunoglobulinas, citoquinas o adipoquinas. Sí que se observó que la variable Ratio_pro_anti_7 y el tiempo de lactancia tienen diferencias en función del tipo de dieta.

Aun así, podemos concluir que no hay diferencias en la composición de la leche en función del tipo de dieta.

Adicionalmente, se quiso ver si los dos grupos de madres que se definieron en el análisis de factores múltiples del Objetivo 1 estaban relacionados con la dieta; es decir, si la mayoría de las madres de un grupo realizaban una de las dietas, y las del otro grupo la otra dieta. Para ello, se colorearon las distintas madres sobre el gráfico de individuos obtenido en función del tipo de dieta que seguían, esperando ver un grupo de un color y el otro del otro. Aun así, tampoco se consiguió ver un patrón que nos permita decir que los grupos que se encontraron estén definidos por el tipo de dieta.

Esta última prueba se realizó también para ver si el hecho de tener o no infecciones podía caracterizar los grupos, de forma que se definió una nueva variable para cada tiempo de infección que en lugar de determinar el número de infecciones, determinaba si había habido o no infección. Aun así, se observó que, en ninguno de los tiempos en los que se observaron las infecciones, no había relación entre los dos grupos de madres y el hecho de haber sufrido o no infecciones.

Caracterización de los dos grupos de madres

Como se pudo observar en el Objetivo 1 mediante la realización del análisis de factores múltiples a dos tiempos, hay dos grupos de madres bien diferenciados y que además tienen comportamientos distintos a lo largo del tiempo. En este apartado, se pretende ver en qué variables se pueden observar diferencias entre los dos grupos y qué características tiene cada uno de ellos.

Para ello, se han separado las madres en dos grupos, en función de la nube de la Figura 2 a la que pertenecieran. Concretamente, se han definido dos inmunotipos; en cada uno de ellos, se incluyen las siguientes madres:

		Madres													
Inmunotipo 1	75	51	103	64	46	53	76	52	40	11	2	105	10		
	50	61	72	90	86	45	29	73	74	35	78	23	56		
	44	98	54	94	71	49	20	85	102	70	84	34	38		
Inmunotipo 2	30	22	57	79	41	33	60	100	47	80	36	1	3		
	104	68	17	6	43	65	48	16	67	93	62	9	99		
	77	28	25	27	12	14	4	42	7	19					

Tabla 2: Distribución de las madres según los dos inmunotipos ⁹

No obstante, y teniendo en cuenta que hay madres que no están concretamente dentro de ninguno de los inmunotipos - y que por lo tanto se pueden considerar como *outliers* - se ha decidido realizar dos veces el mismo estudio: una para los inmunotipos tal y como se han especificado en la tabla anterior, y otro sin tener en cuenta los *outliers*, de forma que, del inmunotipo 1, no se tendrían en cuenta las madres 51 y 75, y del inmunotipo 2, no se tendrían en cuenta las madres 41, 79, 57, 22 y 30.

⁹Elaboración propia.

Análisis bivariante: ¿en qué variables podemos considerar que los inmunotipos se comportan de forma distinta?

Un primer análisis que se ha realizado para caracterizar los dos grupos, es el de ver qué diferencias hay entre ellos, o dicho de otra forma, en qué variables de las que se han incluido en el estudio se puede considerar que los inmunotipos se comportan de forma distinta.

Este análisis se ha realizado teniendo en cuenta las mismas pruebas que se usaron en el caso del análisis bivariante para la comparación de dietas. Dichas pruebas se aplicaron dos veces: una para los datos con valores *outliers*, y otra sin ellos. Los resultados que se han obtenido no son iguales para los dos conjuntos de datos; aunque hay variables que son estadísticamente significativas en ambos casos, no es así en todas. Las tablas que se muestran a continuación presentan un resumen de aquellas variables significativas en cada caso.

	Variable	IC 95 %	p-valor (paramétrico)	p-valor (no paramétrico)
Inmunoglobulinas	Immuno_IgG1_7	(1.132,1.575)	2,83E-19	2,65E-12
	Immuno_IgG_7	(1.196,1.675)	8,01E-18	2,51E-15
	Immuno_IgG_15	(-1.666,-1.138)	2,04E-16	1,17E-14
	Immuno_IgG1_15	(-1.706,-1.145)	2,22E-15	1,26E-11
	Immuno_IgG2_7	(0.882,1.359)	1,10E-13	7,17E-11
	Immuno_IgE_7	(0.004,0.006)	1,72E-12	3,51E-13
	Immuno_IgG2_15	(-1.014,-0.589)	1,08E-10	1,25E-08
	Immuno_IgG4_7	(0.434,0.75)	1,59E-09	4,19E-09
	Immuno_IgG3_7	(0.367,0.677)	3,81E-09	1,44E-09
	Immuno_IgE_15	(-0.003,-0.002)	1,66E-08	7,46E-11
	Immuno_IgG4_15	(-0.668,-0.344)	8,60E-08	3,10E-08
	Immuno_IgG3_15	(-0.801,-0.414)	1,59E-07	1,27E-11
Citoquinas	Cito_IL18_7		0,00	0,00
	Cito_IL22_15		0,00	0,00
	Cito_IL22_7		0,00	0,00
	Cito_IL10_15		0,01	0,00
	Cito_IL12_15		0,01	0,01
Adipoquinas	Leptin_7	(-1.25,-0.369)	0,00	0,00
	Adiponectin_15	(-0.892,-0.228)	0,00	0,00
Vars. calculadas	Immuno_Th1_7	(1.171,1.654)	0	0
	Immuno_Th1_15	(-1.642,-1.11)	0	0
	Immuno_Th2_7	(0.434,0.75)	0	0
	Immuno_Th2_15	(-0.643,-0.317)	0	0
	Ratio_L_A_15	(0.047,0.135)	0	0
	Ratio_L_A_7	(-0.083,-0.013)	0,01	0
	Ratio_Innate_acquired_15	(-2.703,-0.346)	0,01	0,02
	Pro_infl_7	(-1.309,-0.118)	0,02	0,01
	pct_Th1_7	(-0.033,-0.002)	0,03	0,14
Anti_infl_7	(-1.879,-0.057)	0,04	0,02	
Vars. numéricas	Altura 7D	(0.011,0.047)	0	0,01
	Weight 7D	(0.009,0.096)	0,02	0,02
	Altura 15D	(0.003,0.042)	0,02	0,04
	Weight 15D	(0.008,0.099)	0,02	0,03
	Altura 1 MM	(0.001,0.042)	0,04	0,05
Contadores	Infections_0_24M		0,02	
Vars. categóricas	Primipara		0,03	0,02

Tabla 3: Variables con diferencias entre los grupos de madres (con outliers)¹⁰

¹⁰Elaboración propia.

	Variable	IC 95 %	p-valor (paramétrico)	p-valor (no paramétrico)
Inmunoglobulinas	Immuno_IgG_7	(1.321,1.671)	2,80E-24	3,65E-19
	Immuno_IgG1_7	(1.192,1.554)	3,91E-23	2,64E-12
	Immuno_IgG2_7	(1.023,1.36)	1,79E-21	3,69E-12
	Immuno_IgG_15	(-1.552,-1.159)	6,81E-21	1,46E-16
	Immuno_IgG1_15	(-1.576,-1.156)	1,08E-19	6,78E-12
	Immuno_IgE_15	(-0.003,-0.002)	2,33E-16	9,73E-12
	Immuno_IgE_7	(0.004,0.006)	3,51E-15	1,39E-12
	Immuno_IgG3_7	(0.464,0.701)	2,72E-14	7,91E-11
	Immuno_IgG2_15	(-0.982,-0.597)	1,39E-11	4,77E-09
	Immuno_IgG4_7	(0.423,0.709)	8,63E-10	7,32E-09
	Immuno_IgG3_15	(-0.605,-0.347)	4,44E-09	7,93E-11
	Immuno_IgG4_15	(-0.668,-0.34)	5,21E-07	5,75E-08
Citoquinas	Cito_IL22_15		1,33E-05	9,30E-06
	Cito_IL22_7		0,00	0,00
	Cito_IL18_7		0,00	0,03
	Cito_IL10_15		0,01	0,00
	Cito_IL18_15		0,02	5,03E-05
	Cito_IL6_7		0,03	0,02
	Cito_TNF_a_7		0,04	0,00
Adipoquinas	Leptin_7	(-1.224,-0.332)	0,00	0,00
	Adiponectin_15	(-0.732,-0.107)	0,01	0,02
Vars. calculadas	Immuno_Th1_7	(1.297,1.651)	0,00	0,00
	Immuno_Th1_15	(-1.524,-1.124)	0,00	0,00
	Immuno_Th2_7	(0.423,0.709)	0,00	0,00
	Immuno_Th2_15	(-0.638,-0.309)	0,00	0,00
	Ratio_L_A_15	(0.038,0.126)	0,00	0,00
	Ratio_L_A_7	(-0.088,-0.014)	0,01	0,00
	pct_Th1_15	(0.003,0.048)	0,03	0,38
	Pro_infl_7	(-1.206,-0.059)	0,03	0,03
	Ratio_Innate_acquired_15	(-2.883,-0.09)	0,04	0,05
Cito_Th1_7	(0.006,1.123)	0,048	0,00	
Vars. numéricas	Altura 7D	(0.011,0.048)	0,00	0,01
	Weight 7D	(0.004,0.094)	0,03	0,04
	Weight 15D	(0.004,0.1)	0,04	0,06
Vars. categóricas	Primipara		0,03	0,02

Tabla 4: Variables con diferencias entre los grupos de madres (sin outliers)¹¹

Como hemos comentado, hay variables que son significativas en ambos análisis, pero las hay que lo son tan solo en uno de ellos. Por una parte, las inmunoglobulinas que se han detectado como significativas en el estudio con *outliers* también han salido significativas en el estudio que no los tiene en cuenta, igual que pasa con las adipoquinas. En cambio, la citoquina IL12 en el tiempo 15 tan solo es significativa cuando se tienen en cuenta los *outliers*, mientras que la IL18 en el día 15 y la IL6 y la TNF, ambas en el día 7, tan solo lo son cuando se eliminan los *outliers* del estudio. En cuanto a las adipoquinas, ninguna de ellas es significativa en ninguno de los casos, por lo que no hay diferencias entre los inmunotipos de madres en este grupo de variables.

Por otro lado, si observamos los resultados de las variables calculadas vemos como, mientras que el porcentaje de Th1 es significativo en el estudio con *outliers* en el día 7, en el estudio sin estos valores lo es en el día 15. Además, hay dos variables más que tan solo son significativas en uno de los estudios: el Anti_infl en el tiempo 7, que lo es en el estudio con *outliers*, y el Th1 para las citoquinas en el día 7, que lo es en el otro.

¹¹Elaboración propia.

La altura y el peso son variables con significación en ambos estudios, la primera a día 7 y la segunda tanto a día 7 como a día 15. Aun así, si se incluyen los *outliers* en el estudio, la altura también es significativa a día 15 y al mes, y además, en este caso también son significativas las infecciones de los cero a los veinticuatro meses. Por último, el hecho de ser el primer parto de la madre o no (variable primipara) es significativo en ambos casos, lo que significa que los dos inmunotipos no se comportan igual en esta variable.

Random Forest: ¿qué variables determinan la pertenencia a un inmunotipo u a otro?

Para poder caracterizar los inmunotipos, se ha realizado un Random Forest - de nuevo, para los dos conjuntos de datos: uno con *outliers* y otro sin ellos. Con eso lo que se pretende es observar qué variables son importantes a la hora de determinar a que inmunotipo corresponde una madre.

Para ello, se ha separado cada conjunto de datos en dos, de forma que se ha obtenido un conjunto de datos de entrenamiento y uno de testeo para cada estudio. Esto se ha hecho de forma aleatoria, y se han seleccionado 2/3 de los datos para el conjunto de entrenamiento, y el tercio restante para el de testeo. Con los conjuntos de entrenamiento, se han realizado dos estudios mediante el Random Forest. A continuación, usando el conjunto de testeo, se ha comprobado si el modelo obtenido consigue predecir correctamente los datos. Como se ha comentado en el apartado de metodología, se ha hecho una imputación de los valores faltantes usando la media ponderada o las categorías más populares, dependiendo de la clase de la variable, ya que el Random Forest no se puede realizar si hay valores faltantes en alguna de las variables.

En ambos modelos, a partir del vigésimo árbol obtenido, aproximadamente, el error es constante (aunque se han realizado 100 árboles, realmente el error en los ochenta últimos es aproximadamente el mismo), así que los modelos a partir de este punto no mejoran.

Además, en los dos - tanto el modelo que tiene en cuenta los *outliers* como el que no los tiene en cuenta - el porcentaje de aciertos conseguido ha sido muy elevado - de hecho, del 100% al predecir los datos de testeo. Por lo tanto, podemos considerar que hemos conseguido dos modelos que consiguen predecir correctamente a qué inmunotipo pertenece cada madre. Además, con la *cross-validation* que se ha hecho se ha comprobado que la bondad de la predicción se mantiene, por lo que sea cual sea el conjunto de datos de entrenamiento que elijamos, la predicción sigue siendo igual de buena.

Hay dos medidas importantes que se pueden obtener al hacer un Random Forest: el Accuracy(o precisión del modelo) y la medida del índice de Gini. Por un lado, la medida de precisión es una medida del grado en que una variable mejora la precisión del bosque al predecir la clasificación. Los valores más altos significan que la variable mejora la predicción. En un sentido aproximado, se puede interpretar como una muestra de la cantidad de aumento en la precisión de la clasificación que se obtiene al incluir la variable en el modelo. La medida del índice de Gini, por otro lado, es una medida de desorden, y se interpreta de la siguiente forma: a mayor medida, mayor importancia tiene la variable en el modelo creado (ya que valores próximos a cero implican un mayor desorden y valores próximos a uno, un menor desorden). Teniendo eso en cuenta, se ha elaborado una tabla que contiene las diez variables más importantes en cada uno de los modelos, y que por lo tanto, son las que mejor discriminan entre los dos inmunotipos.

Las medidas para cada uno de los modelos que se han efectuado se muestran en los siguientes gráficos:

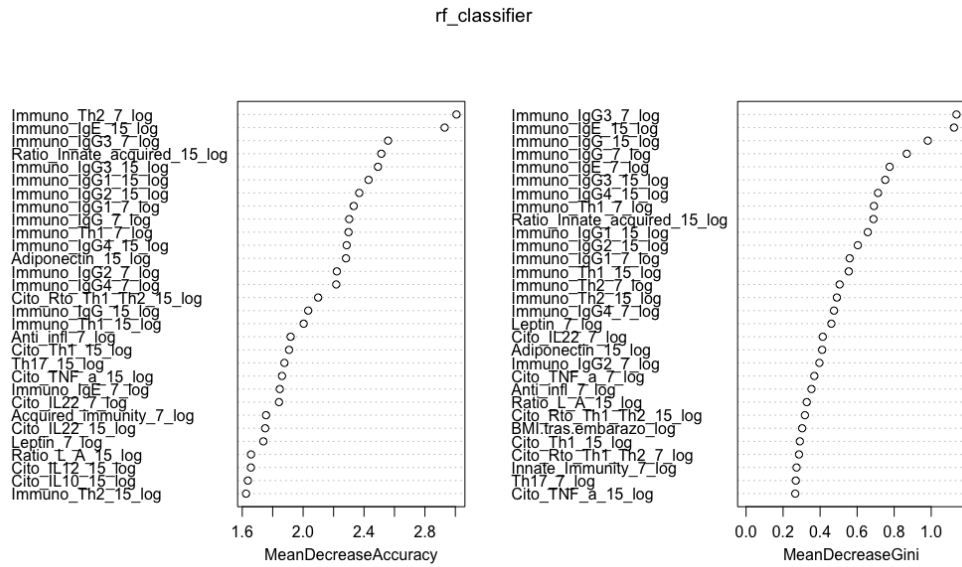


Figura 5: Medidas para el árbol con madres outliers¹²

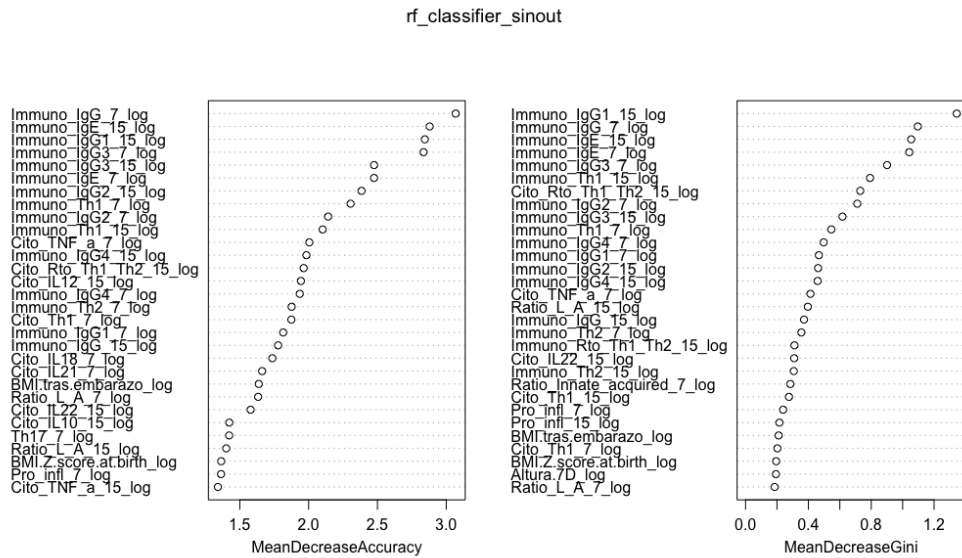


Figura 6: Medidas para el árbol sin madres outliers¹³

Como se puede ver en los gráficos, las variables varían de un modelo a otro, tanto en una medida como en la otra. Por lo tanto, dependiendo de si se deciden incluir o no las madres que parecen alejarse de la nube de puntos, las conclusiones que se obtendrán de este análisis serán distintas, ya que las variables importantes a la hora de determinar a qué grupo pertenece cada madre cambiarán.

Este método nos ha servido para corroborar la caracterización de los dos inmunitipos; los resultados que se han obtenido son los esperados, ya que coinciden con los resultados del análisis de factores múltiples mediante el que se visualizaron los inmunitipos, y dicho análisis se realizó teniendo en cuenta los valores

¹²Elaboración propia.

¹³Elaboración propia.

inmunológicos, que son las variables que han salido con más importancia en el Random Forest. Por lo tanto, no nos interesaban tanto estas variables como las que quedan en segundo plano, que son las de las características de las madres y los bebés, ya que esas no se usaron a la hora de definir los inmunotipos. Es relevante, pues, observar que hay otras variables que ayudan a caracterizar los dos inmunotipos.

V. Presentación de resultados: Dashboard

Como ya se ha comentado, este estudio se ha hecho en colaboración con el departamento de Bioquímica y Fisiología de la facultad de Farmacia y Ciencias de los Alimentos (Universidad de Barcelona), el instituto de Investigación en Nutrición y Seguridad Alimentaria (INSA-UB) y el Instituto de Agroquímica y Tecnología de los Alimentos (IATA-CSIC). Los resultados que se presentan en este trabajo corresponden a los análisis estadísticos que se han realizado para dar forma a los datos con los que están trabajando desde estos departamentos, y poder dar una respuesta estadística a las preguntas que se formulaban.

Con la finalidad de presentar los resultados del estudio de una forma más dinámica y visual a los compañeros, se han elaborado dos *dashboards* - uno correspondiente al primer objetivo del estudio, y otro al segundo - que se han ido presentando a los compañeros y que han servido para compartir los avances del estudio y futuras propuestas. Dichos *dashboards* se han realizado desde el mismo software R, y contienen, en pestañas distintas, los resultados que se han conseguido en cada uno de los análisis realizados.

En concreto, en el dashboard del primer objetivo se muestra, en una primera pestaña, la base de datos completa. En la siguiente pestaña, se puede ver un análisis descriptivo de todas las medidas inmunológicas estudiadas en este objetivo, antes y después de aplicarles la transformación logarítmica mediante la que se ha conseguido normalizar los datos. Por otro lado, se pueden ver los resultados de las pruebas bivanriantes para comparar las medidas del día 7 y del día 15, tanto numéricos como gráficos, por lo que se presenta tanto la tabla con los p-valores de las pruebas realizadas como los boxplots comparativos entre los dos tiempos. Finalmente, en una última pestaña se presentan los resultados de los tres análisis de factores múltiples que se han realizado - el análisis para los dos tiempos, el análisis para el incremento entre tiempos y el análisis para el incremento entre tiempos sin tener en cuenta el grupo de las adipoquinas.

A continuación se adjunta una imagen de muestra de una de las pestañas de este *dashboard*, concretamente la perteneciente al análisis bivanriante de las inmunoglobulinas (aunque no se puede mostrar la página completa por falta de espacio):

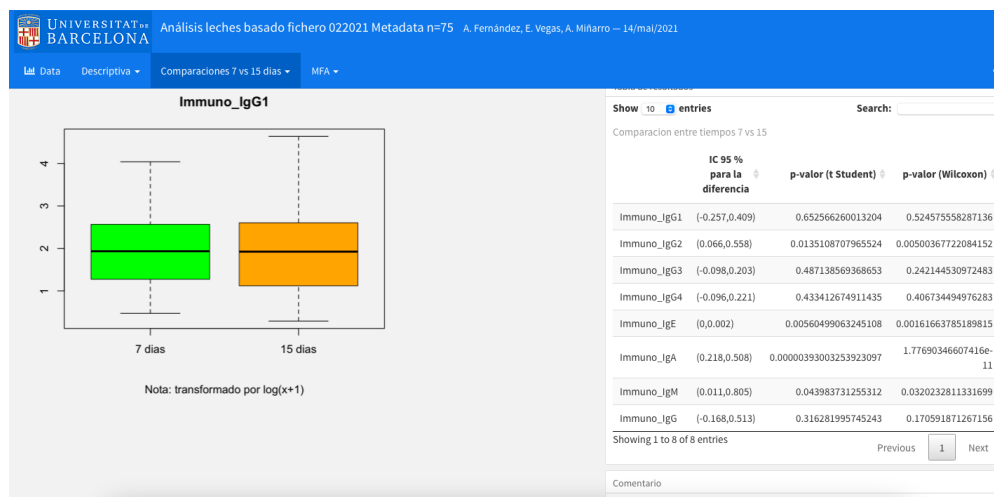


Figura 7: Captura de pantalla de una pestaña del dashboard del primer objetivo¹⁴.

Como se puede observar, se ha seleccionado la pestaña 'Comparaciones 7 vs 15 días'; dentro de este nivel, se desglosan cuatro pestañas adicionales, cada una correspondiente a un grupo de variables (inmunoglobulinas, citoquinas, adipoquinas y variables calculadas) de forma que el usuario puede decidir sobre cuál de ellos quiere ver la comparación. Dicho desglose se encuentra también en el caso de la pestaña 'Descriptiva'.

¹⁴Elaboración propia.

En el caso del MFA, el desglose permite al usuario ver una pequeña descripción de la técnica, así como ver los resultados de los tres análisis realizados (para los dos tiempos, para el incremento entre tiempos y éste último sin incluir las adipocinas). Estos resultados se dividen en distintas secciones: resultados gráficos, tablas de contribuciones (que incluye la tabla con la contribución de cada variable en cada una de las dimensiones y la tabla con la correlación de las mismas con cada dimensión), variables factor map (donde se grafican las variables con una correlación superior a 0.5 sobre las dimensiones) y variables $\cos^2 > 0.75$ (donde se grafican las variables con una correlación superior a 0.75 sobre las dimensiones, en caso de que haya alguna).

Para el segundo objetivo, las dos primeras pestañas contienen la misma información que en el anterior (añadiendo las variables adicionales que se han tenido en cuenta tan solo en este objetivo, correspondientes a las características de madres e hijos). En una tercera pestaña, se presentan los gráficos longitudinales, mediante los que se puede ver la evolución del BMI, el peso y las infecciones a lo largo del tiempo. A continuación, hay los resultados de las pruebas bivariantes para la comparación entre los dos tipos de dieta; igual que en el primer objetivo, dichos resultados se presentan de forma numérica y también gráfica. Seguidamente, se muestra el gráfico de individuos del análisis de factores múltiples realizado en el objetivo anterior, diferenciando las madres según la dieta o según las infecciones en algún momento del tiempo. Finalmente, en las dos siguientes pestañas se encuentra el análisis de caracterización de los dos inmunotipos de madres (tanto teniendo en cuenta todas las madres del estudio, como eliminando aquellas que se pueden considerar *outliers* por no estar bien definidas dentro de un grupo concreto). En ellas se encuentra el mismo estudio bivariate que el que se hizo para las dietas, pero esta vez para la comparación entre los dos inmunotipos, además de los resultados del Random Forest que se ha realizado en cada caso.

Como en el caso del *dashboard* anterior, a continuación se presenta una imagen con una de las pestañas que se pueden encontrar en él:

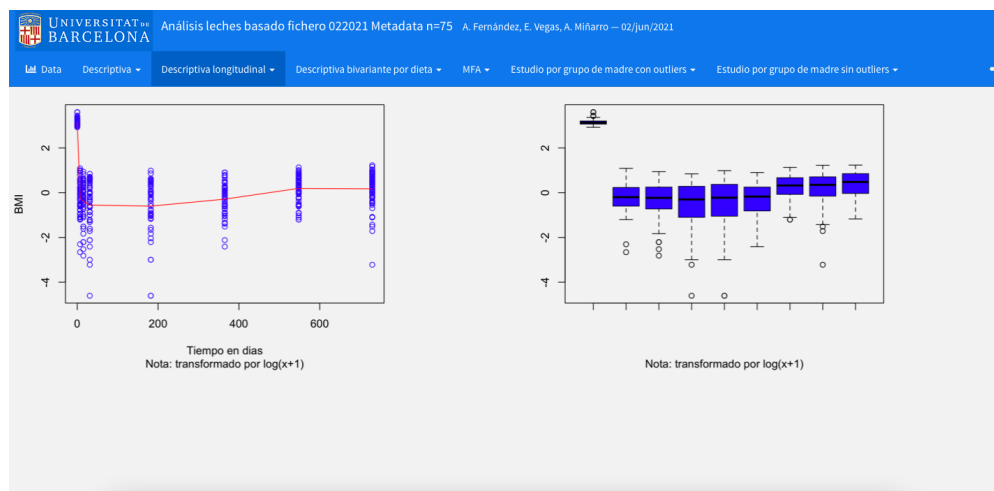


Figura 8: Captura de pantalla de una pestaña del dashboard del segundo objetivo¹⁵.

En la imagen se puede observar la evolución del índice de masa corporal a lo largo del tiempo, tanto en un gráfico que representa todos los individuos en cada tiempo y la línea con la media, como con boxplots en cada tiempo. Como se puede observar, está dentro de la pestaña "Descriptiva longitudinal", donde también se puede encontrar el mismo análisis para el peso y para las infecciones.

Además, las pestañas correspondientes a la descriptiva, al análisis de la dieta, al MFA o los análisis de los inmunotipos tienen, a su vez, varias secciones dentro que el usuario puede seleccionar. En este caso, hay más niveles porque se tienen más tipos de variables - ya que se incluyen las características de la madre y del bebé, que pueden ser numéricas, categóricas, o las que hemos considerado como contadores -.

¹⁵Elaboración propia.

Conclusiones

En este trabajo, se han analizado distintos valores inmunológicos que forman parte de los componentes de la leche materna de 75 pares de madres e hijos, midiéndolos en distintos momentos del tiempo de la etapa de transición, concretamente a los 7 y a los 15 días. Además, se tenía información de distintas características de las madres y de los hijos, como la dieta que siguen, la altura o el peso. Con eso, se quería dar respuesta a dos objetivos: si hay diferencias entre la composición de la leche materna a los 7 días después del parto y a los 15 días después del parto, y qué características son las que influyen en dicha composición.

Para determinar si hay diferencia entre los dos tiempos, se han realizado distintos análisis estadísticos, como pruebas bivariantes para determinar en qué variables se puede considerar que las medidas entre los dos días son distintas, con las que se ha podido observar que hay variables donde sí que se pueden observar estas diferencias. Además se ha realizado un análisis de factores múltiples para caracterizar las madres en los distintos tiempos, con el que se ha visto que hay dos grupos bien diferenciados de madres - que, por lo tanto, tienen valores inmunológicos distintos - que se han denominado como inmunotipos. Se ha visto que las madres de un grupo se comportan de forma distinta a las del otro grupo.

En cuanto a la determinación de qué variables son influyentes a la hora de determinar la composición de la leche, se ha usado como premisa que la dieta podía ser un factor relevante; no obstante, con la realización de pruebas bivariantes se ha comprobado que no es así, y que por lo tanto la dieta por sí sola no tiene relevancia sobre la composición de la leche.

En cambio, cuando se ha estudiado qué variables son distintas entre los dos inmunotipos, definidos gracias al análisis de factores múltiples, sí que se ha observado que hay muchas variables en las que hay diferencias de un inmunotipo a otro. Esto se ha confirmado con la realización de un Random Forest, mediante el cuál hemos corroborado que los valores inmunológicos definen dos grupos de madres bien diferenciados y que, además, las madres de ambos grupos también difieren en otras características.

Por lo tanto, en este trabajo se ha conseguido dar respuesta a la primera pregunta que nos formulábamos, ya que hemos comprobado que hay diferencias en la composición de la leche materna entre los dos tiempos (día 7 y día 15 después del parto). En cuanto al segundo objetivo propuesto, se ha visto que la dieta no es un factor influyente para determinar la composición de la leche materna. Además se han detectado dos inmunotipos distintos que se han podido caracterizar.

No obstante, queda pendiente para un posterior análisis el echo de poder dar una respuesta más específica a cuáles son los factores que influyen en la composición de la leche materna, ya que en el presente estudio, aunque se ha descartado la influencia de la dieta y se han detectado dos grupos diferenciados de madres con diferencias en sus valores inmunológicos y comportamientos distintos, no se ha llegado a dar una respuesta exacta a esa cuestión.

Referencias

- [1] Agarwal S, Karmaus W, Davis S, Gangur V. Immune markers in breast milk and fetal and maternal body fluids: a systematic review of perinatal concentrations. *J Hum Lact.* 2011 May;27(2):171-86. doi: 10.1177/0890334410395761. PMID: 21678611.
- [2] Amat Rodrigo, Joaquín. (Julio de 2017). Test de Wilcoxon Mann Whitney como alternativa al t-test. A: ***Ciencia de Datos***. Recuperado el 02 de abril de 2021. Disponible en: <https://www.cienciadedatos.net/documentos/17_mann-whitney_u_test#Mann_Whitney_Wilcoxon>
- [3] Amat Rodrigo, Joaquín. (Abril de 2018). Tobit Regression: modelos lineales para datos censurados. ***Ciencia de Datos***. Recuperado el 02 de abril de 2021. Disponible en: https://www.cienciadedatos.net/documentos/40_tobit_regression_modelos_lineales_para_datos_censurados
- [4] Amat Rodrigo, Joaquín. (Enero de 2016). T-test. A: ***Ciencia de Datos***. Recuperado el 02 de abril de 2021. Disponible en: <https://www.cienciadedatos.net/documentos/12_t-test>
- [5] Andreas NJ, Kampmann B, Mehring Le-Doare K. Human breast milk: A review on its composition and bioactivity. *Early Hum Dev.* 2015 Nov;91(11):629-35. doi: 10.1016/j.earlhumdev.2015.08.013. Epub 2015 Sep 12. PMID: 26375355.
- [6] Ballard O, Morrow AL. Human milk composition: nutrients and bioactive factors. *Pediatr Clin North Am.* 2013 Feb;60(1):49-74. doi: 10.1016/j.pcl.2012.10.002. PMID: 23178060; PMCID: PMC3586783.
- [7] Bleda Hernández, M.J., Tobías Garcés, A. (7 de enero de 2002). Aplicación de los modelos de regresión tobit en la modelización de variables epidemiológicas censuradas. A: ***Scielo Isciii***. Recuperado el 02 de abril de 2021. Disponible en: <https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112002000200013>
- [8] Cacho NT, Lawrence RM. Innate Immunity and Breast Milk. *Front Immunol.* 2017 May 29;8:584. doi: 10.3389/fimmu.2017.00584. PMID: 28611768; PMCID: PMC5447027.
- [9] Castellote C, Casillas R, Ramírez-Santana C, Pérez-Cano FJ, Castell M, Moretones MG, López-Sabater MC, Franch A. Premature delivery influences the immunological composition of colostrum and transitional and mature human milk. *J Nutr.* 2011 Jun;141(6):1181-7. doi: 10.3945/jn.110.133652. Epub 2011 Apr 20. PMID: 21508211.
- [10] Citoblasto. A: ***Dicciomed***. Recuperado el 22 de abril de 2021. Disponible en: <<https://dicciomed.usal.es/palabra/citoblasto>>
- [11] Citocinas y sus efectos secundarios. A: ***American Cancer Society***. Recuperado el 15 de abril de 2021. Disponible en: <<https://www.cancer.org/es/tratamiento/tratamientos-y-efectos-secundarios/tipos-de-tratamiento/inmunoterapia/citocinas.html>>
- [12] Componentes de la leche materna: ¿Qué contiene la leche materna? A: ***Medela***. Recuperado el 02 de abril de 2021. Disponible en: <<https://www.medela.es/lactancia/viaje-de-las-madres/componentes-de-la-leche-materna>>
- [13] Cortes-Macías E, Selma-Royo M, García-Mantrana I, Calatayud M, González S, Martínez-Costa C, Collado MC. Maternal Diet Shapes the Breast Milk Microbiota Composition and Diversity: Impact of Mode of Delivery and Antibiotic Exposure. *J Nutr.* 2021 Feb 1;151(2):330-340. doi: 10.1093/jn/nxaa310. PMID: 33188413; PMCID: PMC7850106.
- [14] Curran, Rhonda: Immunoglobulina A (IgA). A: ***British Society for immunology***. Recuperado el 15 de abril de 2021. Disponible en: <<https://www.immunology.org/es/public-information/bitesized-immunology/receptors-and-molecules/inmunoglobulina-iga>>
- [15] Dawod B, Marshall JS. Cytokines and Soluble Receptors in Breast Milk as Enhancers of Oral Tolerance Development. *Front Immunol.* 2019 Jan 22;10:16. doi: 10.3389/fimmu.2019.00016. PMID: 30723472; PMCID: PMC6349727.

- [16] Frost BL, Jilling T, Lapin B, Maheshwari A, Caplan MS. Maternal breast milk transforming growth factor-beta and feeding intolerance in preterm infants. *Pediatr Res.* 2014 Oct;76(4):386-93. doi: 10.1038/pr.2014.96. Epub 2014 Jul 4. PMID: 24995914; PMCID: PMC4467901.
- [17] García-Mantrana I, Alcántara C, Selma-Royo M, Boix-Amorós A, Dzidic M, Gimeno-Alcañiz J, Úbeda-Sansano I, Sorribes-Monrabal I, Escuriet R, Gil-Raga F, Parra-Llorca A, Martínez-Costa C, Collado MC; MAMI team. MAMI: a birth cohort focused on maternal-infant microbiota during early life. *BMC Pediatr.* 2019 May 3;19(1):140. doi: 10.1186/s12887-019-1502-y. PMID: 31053102; PMCID: PMC6498642.
- [18] Gavilanes Parra, Sandra, Manjarrez Hernández, Ángel y Cravioto, Alejandro. (Mayo-junio de 2002). Inmunoprotección por leche humana. A: *Revista mexicana de pediatría*. Recuperado el 15 de abril de 2021. Disponible en: <<https://www.medigraphic.com/pdfs/pediat/sp-2002/sp023h.pdf>>
- [19] Inmunoglobulina. A: *Diccionario médico (Clínica Universidad de Navarra)*. Recuperado el 15 de abril de 2021. Disponible en: <<https://www.cun.es/diccionario-medico/terminos/inmunoglobulina>>
- [20] La composición de la leche materna en cada etapa de crecimiento: así se adapta a las necesidades del bebé. A: *Bebés y más*. (Actualizado en 3 de octubre de 2019). Recuperado el 02 de abril de 2021. Disponible en: <<https://www.bebesymas.com/lactancia/composicion-leche-materna-cada-etapa-crecimiento-asi-se-adapta-a-necesidades-bebe>>
- [21] Lactancia materna. A: *Cuidate plus*. Recuperado el 1 de mayo de 2021. Disponible en: <<https://cuidateplus.marca.com/familia/bebe/diccionario/lactancia-materna.html>>
- [22] Métodos paramétricos para la comparación de dos muestras. A: *Fisterra*. (Actualizado en 17 de agosto de 2007). Recuperado el 02 de abril de 2021. Disponible en: <<https://www.fisterra.com/formacion/metodologia-investigacion/metodos-no-parametricos-para-comparacion-dos-muestras/#23193>>
- [23] Pértega Díaz S., Pita Fernández S. (Actualizado en 23 de marzo de 2001). Métodos paramétricos para la comparación de dos medias. t de Student. A: *Fisterra*. Recuperado el 02 de abril de 2021. Disponible en: <https://www.fisterra.com/mbe/investiga/t_student/t_student2.pdf>
- [24] Pruebas no paramétricas: definición y tipos. A: *La mente es maravillosa*. (28 de enero de 2019). Recuperado el 02 de abril de 2021. Disponible en: <<https://lamenteesmaravillosa.com/pruebas-no-parametricas-definicion-y-tipos/>>
- [25] ¿Qué es la leche de transición? A: *Medela*. Recuperado el 02 de abril de 2021. Disponible en: <<https://www.medela.es/lactancia/viaje-de-las-madres/leche-de-transicion>>
- [26] ¿Qué son las pruebas paramétricas? A: *QuestionPro* Recuperado el 02 de abril de 2021. Disponible en: <<https://www.questionpro.com/blog/es/pruebas-parametricas/>>

Anexos

Anexo A. Resultados del Objetivo 1

A.1. Análisis descriptivo inicial

	% de 0's		% de 0's
Cito_IL13_7	98,7	Cito_IL1b_15	73,3
Cito_IL27_15	98,7	Cito_IL12_7	66,7
Cito_IL4_15	98,7	Cito_IL1b_7	54,7
Cito_GM-CSF_15	98,7	Cito_IL10_7	54,7
Cito_IL27_7	97,3	Cito_IL6_15	54,7
Cito_IL13_15	97,3	Cito_IL10_15	54,7
Cito_IL2_15	96,0	Cito_IL12_15	54,7
Cito_IL23_15	96,0	Acquired_immunity_15	52,0
Cito_IL4_7	93,3	Acquired_immunity_7	48,0
Cito_GM-CSF_7	93,3	Anti_infl_15	41,3
Cito_IL5_15	93,3	Cito_IL6_7	38,7
Cito_IL23_7	92,0	Anti_infl_7	32,0
Cito_IL17_15	92,0	Cito_IL22_7	25,3
Hematopoyetic_factors_15	92,0	Cito_IL22_15	25,3
Cito_Th2_15	89,3	Cito_IL21_7	14,7
Cito_IL9_15	88,0	Cito_TNF-a_15	14,7
Th9_15	88,0	Cito_IL21_15	14,7
Cito_IL2_7	86,7	Cito_Th1_15	13,3
Cito_IL5_7	86,7	Cito_TNF-a_7	12,0
Cito_IL9_7	85,3	Cito_Th1_7	10,7
Th9_7	85,3	Th17_15	9,3
Cito_IL17_7	84,0	Innate_Immunity_15	9,3
Hematopoyetic_factors_7	84,0	Cito_IL18_15	8,0
Cito_Ifn-g_7	80,0	Cito_IL18_7	5,3
Cito_Ifn-g_15	78,7	Th17_7	5,3
Cito_Th2_7	77,3	Innate_Immunity_7	5,3

Tabla 5: Porcentaje de ceros por variable

	N	Media	Desv. Est.	Mediana	Mínimo	Máximo	Error Est.
Immuno_IgG1_7	75	9,1	9,5	5,9	0,6	55,7	1,1
Immuno_IgG2_7	75	6,2	6,0	4,7	0,6	30,1	0,7
Immuno_IgG3_7	75	1,0	1,1	0,7	0,0	6,2	0,1
Immuno_IgG4_7	75	0,8	1,1	0,3	0,0	5,4	0,1
Immuno_IgE_7	75	0,0	0,0	0,0	0,0	0,0	0,0
Immuno_IgA_7	75	1176,8	479,0	1137,0	31,2	2470,0	55,3
Immuno_IgM_7	75	157,6	290,5	78,7	3,2	1341,0	33,5
Immuno_IgG_7	75	17,1	16,1	15,6	1,9	86,1	1,9
Immuno_IgG1_15	75	9,9	15,6	5,8	0,3	102,0	1,8
Immuno_IgG2_15	75	3,8	3,0	2,9	0,2	13,2	0,3
Immuno_IgG3_15	75	1,0	2,1	0,5	0,0	14,4	0,2
Immuno_IgG4_15	75	0,7	0,9	0,2	0,0	3,6	0,1
Immuno_IgE_15	75	0,0	0,0	0,0	0,0	0,0	0,0
Immuno_IgA_15	75	803,9	384,9	725,5	31,2	2500,0	44,4
Immuno_IgM_15	75	147,4	321,1	50,8	1,9	1357,2	37,1
Immuno_IgG_15	75	15,4	20,1	10,3	1,3	128,3	2,3
Cito_IL27_7	75	0,3	1,9	0,0	0,0	14,9	0,2
Cito_IL1b_7	75	5,4	20,6	0,0	0,0	168,4	2,4
Cito_IL2_7	75	1,5	4,9	0,0	0,0	32,4	0,6
Cito_IL4_7	75	0,2	1,2	0,0	0,0	7,4	0,1
Cito_IL5_7	75	0,7	2,2	0,0	0,0	15,7	0,3
Cito_IL6_7	75	47,8	132,7	5,1	0,0	1083,9	15,3
Cito_IL10_7	75	1,2	5,2	0,0	0,0	43,2	0,6
Cito_IL12_7	75	0,4	2,7	0,0	0,0	23,2	0,3
Cito_IL13_7	75	0,0	0,4	0,0	0,0	3,2	0,0
Cito_IL17_7	75	0,6	2,1	0,0	0,0	14,5	0,2
Cito_Ifn-g_7	75	0,5	1,9	0,0	0,0	12,1	0,2
Cito_GM-CSF_7	75	1,0	5,9	0,0	0,0	48,6	0,7
Cito_TNF-a_7	75	2,6	3,8	1,8	0,0	21,4	0,4
Cito_IL9_7	75	0,6	2,4	0,0	0,0	13,1	0,3
Cito_IL23_7	75	0,5	1,9	0,0	0,0	10,0	0,2
Cito_IL18_7	75	16,5	26,5	8,1	0,0	187,7	3,1
Cito_IL21_7	75	11,4	30,7	2,4	0,0	193,3	3,5
Cito_IL22_7	75	8,0	9,5	4,1	0,0	47,2	1,1
Cito_IL27_15	75	1,8	15,7	0,0	0,0	136,2	1,8
Cito_IL1b_15	75	8,2	55,1	0,0	0,0	475,2	6,4
Cito_IL2_15	75	0,1	0,8	0,0	0,0	5,0	0,1
Cito_IL4_15	75	0,0	0,1	0,0	0,0	0,6	0,0
Cito_IL5_15	75	0,7	3,9	0,0	0,0	28,1	0,5
Cito_IL6_15	75	30,2	102,4	0,0	0,0	806,6	11,8
Cito_IL10_15	75	0,6	1,8	0,0	0,0	10,9	0,2
Cito_IL12_15	75	0,1	0,2	0,0	0,0	0,6	0,0
Cito_IL13_15	75	0,0	0,2	0,0	0,0	1,3	0,0
Cito_IL17_15	75	0,2	0,7	0,0	0,0	3,6	0,1
Cito_Ifn-g_15	75	1,1	3,6	0,0	0,0	26,3	0,4
Cito_GM-CSF_15	75	0,0	0,4	0,0	0,0	3,4	0,0
Cito_TNF-a_15	75	1,7	1,8	1,7	0,0	9,1	0,2
Cito_IL9_15	75	0,1	0,5	0,0	0,0	3,8	0,1
Cito_IL23_15	75	0,1	0,7	0,0	0,0	3,4	0,1
Cito_IL18_15	75	26,3	50,9	13,2	0,0	388,6	5,9
Cito_IL21_15	75	5,7	11,3	2,2	0,0	72,6	1,3
Cito_IL22_15	75	7,2	10,5	4,3	0,0	75,2	1,2
Leptin_7	75	589,2	588,0	343,5	17,5	2767,9	67,9
Adiponectin_7	75	12971,0	9102,6	10703,6	1068,1	48170,5	1051,1
Leptin_15	75	644,0	675,4	409,5	7,8	2841,9	78,0
Adiponectin_15	75	13857,9	15650,2	10813,9	811,6	127358,5	1807,1

Tabla 6: Estadísticos iniciales (immunoglobulinas, citoquinas y adipocinas)

	N	Media	Desv. Est.	Mediana	Mínimo	Máximo	Error Est.
Ratio_L_A_7	75	0,1	0,1	0,0	0,0	0,5	0,0
Pro_infl_7	75	75,8	145,1	26,8	1,7	1100,6	16,8
Anti_infl_7	75	50,5	137,5	5,7	0,0	1127,1	15,9
Ratio_pro_anti_7	51	38,1	171,0	1,5	1,0	1158,2	23,9
Cito_Th1_7	75	10,7	24,4	2,0	0,0	172,5	2,8
Cito_Th2_7	75	2,5	6,8	0,0	0,0	38,4	0,8
Cito_Rto_Th1_Th2_7	17	2,1	1,4	2,1	0,0	4,6	0,3
Th17_7	75	73,6	138,2	29,3	0,0	1087,7	16,0
Th9_7	75	0,6	2,4	0,0	0,0	13,1	0,3
Hematopoyetic_factors_7	75	1,7	7,0	0,0	0,0	51,1	0,8
Innate_Immunity_7	75	56,7	137,4	9,2	0,0	1094,9	15,9
Acquired_immunity_7	75	3,6	9,5	0,0	0,0	51,9	1,1
Ratio_Innate_acquired_7	39	191,4	384,3	34,6	0,0	2144,8	61,5
Immuno_Th1_7	75	16,3	15,3	14,4	1,7	82,1	1,8
Immuno_Th2_7	75	0,8	1,1	0,3	0,0	5,4	0,1
Immuno_Rto_Th1_Th2_7	75	38,1	34,9	25,5	4,2	152,2	4,0
pct_Th1_7	75	95,5	3,3	96,2	80,8	99,3	0,4
pct_Th2_7	75	4,5	3,3	3,8	0,7	19,2	0,4
pct_Rto_Th1_Th2_7	75	38,1	34,9	25,5	4,2	152,2	4,0
Ratio_L_A_15	75	0,1	0,1	0,0	0,0	0,8	0,0
Pro_infl_15	75	68,0	159,6	20,2	0,2	1029,8	18,4
Anti_infl_15	75	30,9	103,2	0,1	0,0	808,8	11,9
Ratio_pro_anti_15	44	132,6	452,8	2,7	1,0	2542,1	68,3
Cito_Th1_15	75	13,1	58,2	1,8	0,0	485,7	6,7
Cito_Th2_15	75	0,9	4,0	0,0	0,0	28,1	0,5
Cito_Rto_Th1_Th2_15	8	7,2	9,8	1,6	0,4	25,5	3,5
Th17_15	75	51,6	129,9	14,9	0,0	867,6	15,0
Th9_15	75	0,1	0,5	0,0	0,0	3,8	0,1
Hematopoyetic_factors_15	75	0,8	3,9	0,0	0,0	28,1	0,5
Innate_Immunity_15	75	41,3	130,7	3,7	0,0	859,3	15,1
Acquired_immunity_15	75	1,0	2,5	0,0	0,0	14,4	0,3
Ratio_Innate_acquired_15	36	133,3	246,6	47,0	0,2	1247,3	41,1
Immuno_Th1_15	75	14,7	19,7	9,9	1,2	126,7	2,3
Immuno_Th2_15	75	0,7	0,9	0,2	0,0	3,6	0,1
Immuno_Rto_Th1_Th2_15	75	44,2	48,8	27,6	4,5	245,4	5,6
pct_Th1_15	75	95,3	4,0	96,5	81,8	99,6	0,5
pct_Th2_15	75	4,7	4,0	3,5	0,4	18,2	0,5
pct_Rto_Th1_Th2_15	75	44,2	48,8	27,6	4,5	245,4	5,6

Tabla 7: Estadísticos iniciales (variables calculadas)

	N	Media	Desv. Est.	Mediana	Mínimo	Máximo	Error Est.
Immuno_IgG1_7_log	75	2,0	0,8	1,9	0,5	4,0	0,1
Immuno_IgG2_7_log	75	1,7	0,8	1,7	0,5	3,4	0,1
Immuno_IgG3_7_log	75	0,6	0,4	0,5	0,0	2,0	0,0
Immuno_IgG4_7_log	75	0,5	0,5	0,3	0,0	1,9	0,1
Immuno_IgE_7_log	75	0,0	0,0	0,0	0,0	0,0	0,0
Immuno_IgA_7_log	75	6,9	0,7	7,0	3,5	7,8	0,1
Immuno_IgM_7_log	75	4,3	1,1	4,4	1,4	7,2	0,1
Immuno_IgG_7_log	75	2,5	0,9	2,8	1,1	4,5	0,1
Immuno_IgG1_15_log	75	1,9	0,9	1,9	0,3	4,6	0,1
Immuno_IgG2_15_log	75	1,4	0,6	1,4	0,2	2,7	0,1
Immuno_IgG3_15_log	75	0,5	0,5	0,4	0,0	2,7	0,1
Immuno_IgG4_15_log	75	0,4	0,4	0,2	0,0	1,5	0,0
Immuno_IgE_15_log	75	0,0	0,0	0,0	0,0	0,0	0,0
Immuno_IgA_15_log	75	6,6	0,5	6,6	3,5	7,8	0,1
Immuno_IgM_15_log	75	3,9	1,3	3,9	1,1	7,2	0,2
Immuno_IgG_15_log	75	2,4	0,9	2,4	0,8	4,9	0,1
Cito_IL27_7_log	75	0,1	0,4	0,0	0,0	2,8	0,0
Cito_IL1b_7_log	75	0,7	1,1	0,0	0,0	5,1	0,1
Cito_IL2_7_log	75	0,3	0,8	0,0	0,0	3,5	0,1
Cito_IL4_7_log	75	0,1	0,4	0,0	0,0	2,1	0,0
Cito_IL5_7_log	75	0,2	0,6	0,0	0,0	2,8	0,1
Cito_IL6_7_log	75	2,1	2,0	1,8	0,0	7,0	0,2
Cito_IL10_7_log	75	0,3	0,7	0,0	0,0	3,8	0,1
Cito_IL12_7_log	75	0,1	0,4	0,0	0,0	3,2	0,0
Cito_IL13_7_log	75	0,0	0,2	0,0	0,0	1,4	0,0
Cito_IL17_7_log	75	0,2	0,5	0,0	0,0	2,7	0,1
Cito_Ifn.g_7_log	75	0,2	0,5	0,0	0,0	2,6	0,1
Cito_GM.CSF_7_log	75	0,2	0,6	0,0	0,0	3,9	0,1
Cito_TNF.a_7_log	75	0,9	0,8	1,0	0,0	3,1	0,1
Cito_IL9_7_log	75	0,2	0,6	0,0	0,0	2,7	0,1
Cito_IL23_7_log	75	0,2	0,5	0,0	0,0	2,4	0,1
Cito_IL18_7_log	75	2,2	1,2	2,2	0,0	5,2	0,1
Cito_IL21_7_log	75	1,5	1,2	1,2	0,0	5,3	0,1
Cito_IL22_7_log	75	1,6	1,2	1,6	0,0	3,9	0,1
Cito_IL27_15_log	75	0,1	0,6	0,0	0,0	4,9	0,1
Cito_IL1b_15_log	75	0,4	1,1	0,0	0,0	6,2	0,1
Cito_IL2_15_log	75	0,1	0,3	0,0	0,0	1,8	0,0
Cito_IL4_15_log	75	0,0	0,1	0,0	0,0	0,5	0,0
Cito_IL5_15_log	75	0,1	0,6	0,0	0,0	3,4	0,1
Cito_IL6_15_log	75	1,4	1,8	0,0	0,0	6,7	0,2
Cito_IL10_15_log	75	0,3	0,5	0,0	0,0	2,5	0,1
Cito_IL12_15_log	75	0,1	0,1	0,0	0,0	0,5	0,0
Cito_IL13_15_log	75	0,0	0,1	0,0	0,0	0,8	0,0
Cito_IL17_15_log	75	0,1	0,3	0,0	0,0	1,5	0,0
Cito_Ifn.g_15_log	75	0,3	0,7	0,0	0,0	3,3	0,1
Cito_GM.CSF_15_log	75	0,0	0,2	0,0	0,0	1,5	0,0
Cito_TNF.a_15_log	75	0,8	0,6	1,0	0,0	2,3	0,1
Cito_IL9_15_log	75	0,1	0,2	0,0	0,0	1,6	0,0
Cito_IL23_15_log	75	0,1	0,3	0,0	0,0	1,5	0,0
Cito_IL18_15_log	75	2,5	1,3	2,7	0,0	6,0	0,2
Cito_IL21_15_log	75	1,3	1,0	1,2	0,0	4,3	0,1
Cito_IL22_15_log	75	1,5	1,1	1,7	0,0	4,3	0,1
Leptin_7_log	75	5,9	1,0	5,8	2,9	7,9	0,1
Adiponectin_7_log	75	9,2	0,8	9,3	7,0	10,8	0,1
Leptin_15_log	75	5,9	1,2	6,0	2,2	8,0	0,1
Adiponectin_15_log	75	9,2	0,8	9,3	6,7	11,8	0,1

Tabla 8: Estadísticos después de la transformación (immunoglobulinas, citoquina y adipocinas)

	N	Media	Desv. Est.	Mediana	Mínimo	Máximo	Error Est.
Ratio.L.A_7_log	75	0,1	0,1	0,0	0,0	0,4	0,0
Pro_infl_7_log	75	3,4	1,3	3,3	1,0	7,0	0,2
Anti_infl_7_log	75	2,2	2,0	1,9	0,0	7,0	0,2
Ratio_pro_anti_7_log	51	1,5	1,4	0,9	0,7	7,1	0,2
Cito_Th1_7_log	75	1,5	1,3	1,1	0,0	5,2	0,1
Cito_Th2_7_log	75	0,5	1,0	0,0	0,0	3,7	0,1
Cito_Rto_Th1_Th2_7_log	17	1,0	0,5	1,1	0,0	1,7	0,1
Th17_7_log	75	3,2	1,6	3,4	0,0	7,0	0,2
Th9_7_log	75	0,2	0,6	0,0	0,0	2,7	0,1
Hematopoyetic_factors_7_log	75	0,3	0,8	0,0	0,0	4,0	0,1
Innate_Immunity_7_log	75	2,6	1,8	2,3	0,0	7,0	0,2
Acquired_immunity_7_log	75	0,6	1,1	0,0	0,0	4,0	0,1
Ratio_Innate_acquired_7_log	39	3,6	2,1	3,6	0,0	7,7	0,3
Immuno_Th1_7_log	75	2,5	0,9	2,7	1,0	4,4	0,1
Immuno_Th2_7_log	75	0,5	0,5	0,3	0,0	1,9	0,1
Immuno_Rto_Th1_Th2_7_log	75	3,4	0,8	3,3	1,7	5,0	0,1
pct_Th1_7_log	75	4,6	0,0	4,6	4,4	4,6	0,0
pct_Th2_7_log	75	1,6	0,6	1,6	0,5	3,0	0,1
pct_Rto_Th1_Th2_7_log	75	3,4	0,8	3,3	1,7	5,0	0,1
Ratio.L.A_15_log	75	0,1	0,1	0,0	0,0	0,6	0,0
Pro_infl_15_log	75	3,1	1,4	3,1	0,2	6,9	0,2
Anti_infl_15_log	75	1,4	1,8	0,1	0,0	6,7	0,2
Ratio_pro_anti_15_log	44	2,1	1,9	1,3	0,7	7,8	0,3
Cito_Th1_15_log	75	1,3	1,2	1,0	0,0	6,2	0,1
Cito_Th2_15_log	75	0,2	0,6	0,0	0,0	3,4	0,1
Cito_Rto_Th1_Th2_15_log	8	1,5	1,1	0,9	0,3	3,3	0,4
Th17_15_log	75	2,8	1,5	2,8	0,0	6,8	0,2
Th9_15_log	75	0,1	0,2	0,0	0,0	1,6	0,0
Hematopoyetic_factors_15_log	75	0,2	0,6	0,0	0,0	3,4	0,1
Innate_Immunity_15_log	75	1,9	1,7	1,6	0,0	6,8	0,2
Acquired_immunity_15_log	75	0,4	0,7	0,0	0,0	2,7	0,1
Ratio_Innate_acquired_15_log	36	3,7	1,7	3,9	0,2	7,1	0,3
Immuno_Th1_15_log	75	2,3	0,9	2,4	0,8	4,9	0,1
Immuno_Th2_15_log	75	0,4	0,4	0,2	0,0	1,5	0,0
Immuno_Rto_Th1_Th2_15_log	75	3,4	0,9	3,4	1,7	5,5	0,1
pct_Th1_15_log	75	4,6	0,0	4,6	4,4	4,6	0,0
pct_Th2_15_log	75	1,5	0,6	1,5	0,3	3,0	0,1
pct_Rto_Th1_Th2_15_log	75	3,4	0,9	3,4	1,7	5,5	0,1

Tabla 9: Estadísticos después de la transformación (variables calculadas)

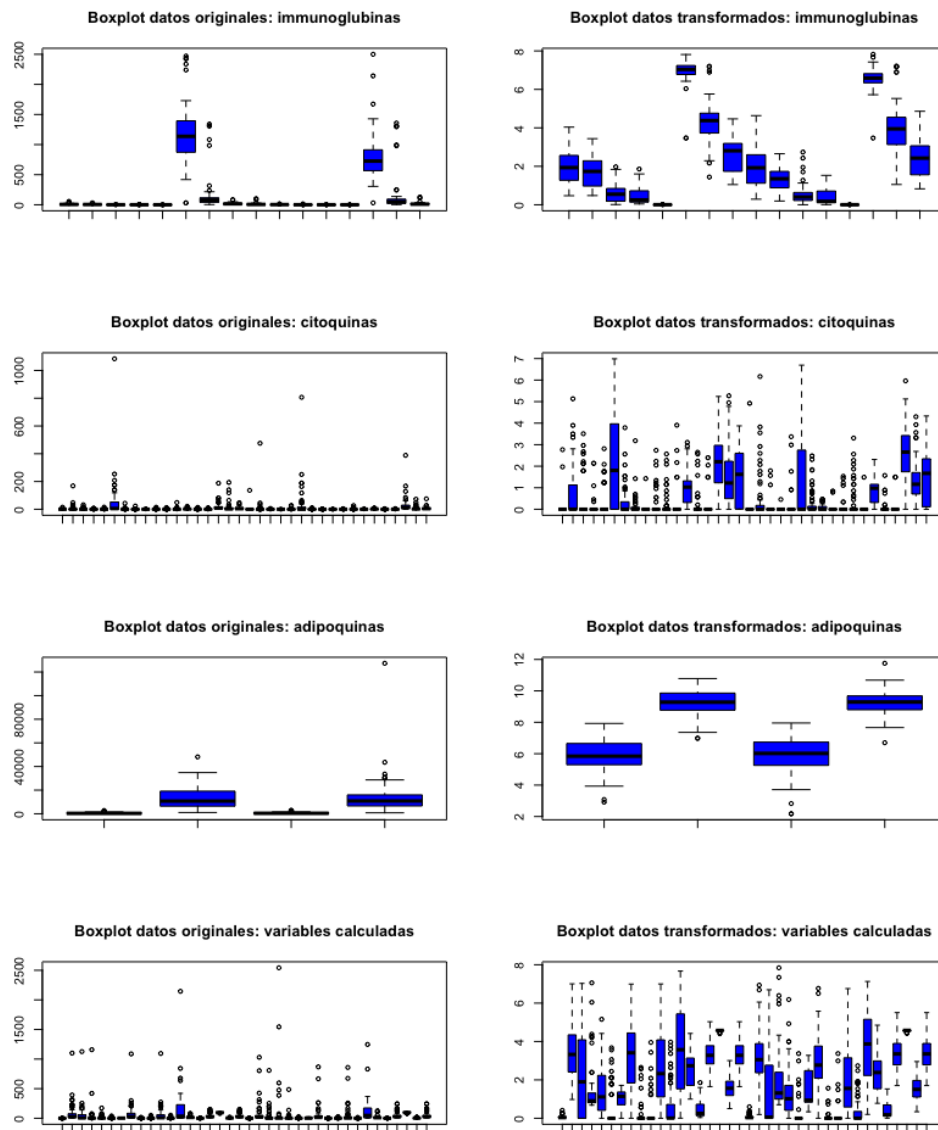


Figura 9: Boxplots múltiples: Comparación entre antes y después de la transformación

A.2. Comparación entre los dos tiempos

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Immuno_IgA	(0.218,0.508)	3,93E-06	1,78E-11
Immuno_IgE	(0,0.002)	0,006	0,002
Immuno_IgG2	(0.066,0.558)	0,014	0,005
Immuno_IgM	(0.011,0.805)	0,044	0,032
Immuno_IgG	(-0.168,0.513)	0,316	0,171
Immuno_IgG4	(-0.096,0.221)	0,433	0,407
Immuno_IgG3	(-0.098,0.203)	0,487	0,242
Immuno_IgG1	(-0.257,0.409)	0,653	0,525

Tabla 10: Resultados de las pruebas para la comparación entre los dos tiempos (immunoglobulinas)

	Q1 y Q3 para la diferencia (15-7)	p-valor (modelo tobit)	p-valor (Wilcoxon)
Cito_Ifn.g	0	0,003	0,216
Cito_IL1b	(-0.855,0)	0,004	0,015
Cito_IL9	0	0,005	0,239
Cito_IL6	(-2.491,0.253)	0,006	0,009
Cito_IL10	(-0.142,0.04)	0,006	0,412
Cito_IL18	(-0.465,1.307)	0,009	0,030
Cito_IL17	0	0,019	0,163
Cito_IL5	0	0,048	0,379
Cito_IL12	(0,0.095)	0,056	0,584
Cito_IL23	0	0,086	0,159
Cito_IL2	0	0,154	0,005
Cito_IL13	0	0,155	1,000
Cito_TNF.a	(-0.98,0.815)	0,299	0,230
Cito_GM.CSF	0	0,308	0,090
Cito_IL27	0	0,308	1,000
Cito_IL4	0	0,308	0,114
Cito_IL22	(-1.27,1.309)	0,491	0,942
Cito_IL21	(-0.889,0.617)	0,756	0,198

Tabla 11: Resultados de las pruebas para la comparación entre los dos tiempos (citoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Adiponectin	(-0.182,0.109)	0,622	0,905
Leptin	(-0.224,0.233)	0,966	0,810

Tabla 12: Resultados de las pruebas para la comparación entre los dos tiempos (adipoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Anti_infl	(0.239,1.293)	0,005	0,006
Innate_Immunity	(0.209,1.158)	0,005	0,004
Cito_Th2	(0.039,0.555)	0,025	0,026
Acquired_immunity	(0.029,0.563)	0,030	0,066
Th17	(-0.025,0.876)	0,064	0,024
Th9	(-0.025,0.27)	0,103	0,239
Pro_infl	(-0.067,0.696)	0,105	0,230
Hematopoyetic_factors	(-0.068,0.39)	0,165	0,184
Ratio_pro_anti	(-1.281,0.33)	0,238	0,068
Cito_Th1	(-0.18,0.588)	0,294	0,174
Immuno_Th1	(-0.161,0.508)	0,305	0,158
Immuno_Th2	(-0.082,0.232)	0,344	0,306
Ratio_L_A	(-0.033,0.017)	0,510	0,710
pct_Th1	(-0.009,0.014)	0,675	0,872
pct_Rto_Th1_Th2	(-0.255,0.17)	0,692	0,581
Immuno_Rto_Th1_Th2	(-0.255,0.17)	0,692	0,581
Cito_Rto_Th1_Th2	(-3.816,3.234)	0,756	1,000
pct_Th2	(-0.141,0.171)	0,851	0,714
Ratio_Innate_acquired	(-1.1,1.248)	0,896	0,829

Tabla 13: Resultados de las pruebas para la comparación entre los dos tiempos (variables calculadas)

A.3. MFA para los dos tiempos

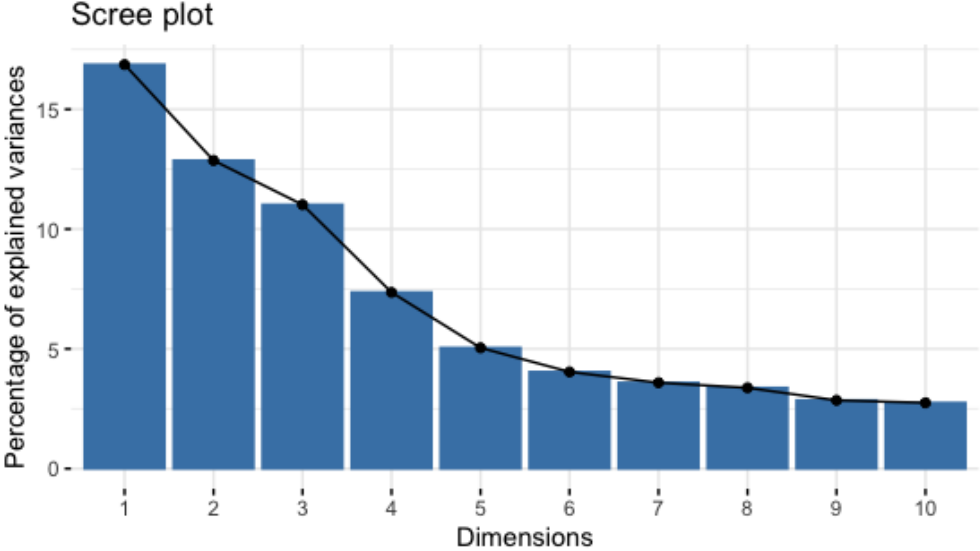


Figura 10: MFA para los dos tiempos: Porcentajes de variabilidad explicados por cada dimensión

	Dim.1	Dim.2	Dim.3	Dim.4
Immuno_IgG_15_log	7,11	0,01	3,32	0,43
Immuno_IgG1_15_log	7,00	0,00	3,21	0,19
Immuno_IgG3_15_log	5,79	0,11	2,77	0,19
Immuno_IgE_15_log	5,64	0,55	2,45	0,86
Cito_Ifn.g_7_log	5,56	0,94	1,04	1,47
Immuno_IgG2_15_log	5,17	0,00	2,68	1,97
Immuno_IgG4_15_log	4,91	0,04	0,73	1,14
Cito_IL18_7_log	3,46	0,04	0,05	0,87
Cito_IL6_7_log	3,06	2,12	0,00	1,30
Cito_IL22_7_log	2,99	0,04	0,40	3,64
Cito_IL17_7_log	2,74	0,00	8,63	0,06
Immuno_IgG1_7_log	2,65	8,25	0,00	0,00
Immuno_IgG_7_log	2,64	8,96	0,02	0,03
Cito_IL5_7_log	2,56	0,34	2,04	0,27
Leptin_7_log	2,55	0,57	0,32	5,04
Cito_IL1b_7_log	2,40	3,60	0,45	0,02
Adiponectin_15_log	2,30	0,03	1,53	0,07
Cito_IL10_7_log	2,14	2,83	0,98	0,43
Cito_IL2_15_log	2,03	1,56	1,53	1,52
Immuno_IgG2_7_log	2,03	8,92	0,09	0,06
Immuno_IgE_7_log	2,02	8,17	0,07	0,17
Cito_GM.CSF_7_log	1,96	0,00	9,47	0,96
Cito_IL6_15_log	1,92	1,91	2,99	1,81
Cito_IL1b_15_log	1,82	0,59	2,07	3,57
Immuno_IgG4_7_log	1,68	7,32	0,05	0,01
Cito_IL2_7_log	1,49	0,12	8,82	0,02
Cito_IL23_15_log	1,46	0,09	4,32	1,94
Immuno_IgG3_7_log	1,41	6,97	0,14	0,13
Cito_IL10_15_log	1,40	2,03	2,23	0,55
Cito_IL9_7_log	1,26	0,03	5,08	0,01
Cito_IL13_7_log	1,24	0,00	6,46	2,41
Cito_TNF.a_15_log	1,10	0,00	1,40	7,98
Cito_IL13_15_log	0,89	3,08	0,90	0,46
Cito_IL12_15_log	0,89	0,36	0,11	6,30
Cito_IL27_7_log	0,76	0,10	8,23	1,73
Cito_TNF.a_7_log	0,74	5,41	2,85	0,66
Cito_IL22_15_log	0,53	3,98	0,25	0,31
Cito_IL23_7_log	0,49	0,30	1,73	0,64
Cito_IL17_15_log	0,43	1,64	1,16	0,52
Cito_IL12_7_log	0,32	0,01	2,71	0,81
Immuno_IgA_15_log	0,28	1,99	0,00	2,39
Cito_Ifn.g_15_log	0,25	1,91	2,76	8,68
Cito_GM.CSF_15_log	0,19	0,02	0,00	0,51
Cito_IL9_15_log	0,14	0,03	0,04	0,01
Cito_IL21_15_log	0,13	1,48	0,00	2,48
Cito_IL27_15_log	0,12	0,03	0,00	0,01
Cito_IL18_15_log	0,11	2,72	0,67	8,17
Immuno_IgA_7_log	0,10	0,53	0,01	1,09
Cito_IL4_7_log	0,04	2,08	0,42	0,74
Cito_IL5_15_log	0,03	0,00	0,00	11,51
Cito_IL21_7_log	0,02	1,37	0,97	4,47
Adiponectin_7_log	0,02	0,57	1,30	0,03
Cito_IL4_15_log	0,01	0,22	0,45	1,54
Immuno_IgM_7_log	0,00	1,34	0,09	0,00
Leptin_15_log	0,00	3,57	0,00	3,28
Immuno_IgM_15_log	0,00	1,14	0,02	4,55

Tabla 14: MFA para los dos tiempos: Contribución de cada variable en cada una de las dimensiones

	Dim.1	Dim.2	Dim.3	Dim.4
Cito_IL27_7_log	0,27	0,09	0,71	0,27
Cito_IL1b_7_log	0,48	0,51	0,17	-0,03
Cito_IL2_7_log	0,38	0,09	0,74	0,03
Cito_IL4_7_log	0,07	0,39	0,16	-0,18
Cito_IL5_7_log	0,49	0,16	0,36	0,11
Cito_IL6_7_log	0,54	0,39	0,02	-0,23
Cito_IL10_7_log	0,45	0,45	0,25	-0,13
Cito_IL12_7_log	0,17	-0,03	0,41	0,18
Cito_IL13_7_log	0,34	0,01	0,63	0,32
Cito_IL17_7_log	0,51	0,01	0,73	0,05
Cito_Ifn.g_7_log	0,73	0,26	0,25	0,25
Cito_GM.CSF_7_log	0,43	0,01	0,77	0,20
Cito_TNF.a_7_log	0,27	0,63	0,42	-0,16
Cito_IL9_7_log	0,35	0,04	0,56	0,02
Cito_IL23_7_log	-0,22	0,15	0,33	-0,16
Cito_IL18_7_log	0,57	0,06	-0,06	0,19
Cito_IL21_7_log	0,04	0,32	0,25	-0,43
Cito_IL22_7_log	0,53	-0,05	0,16	-0,39
Immuno_IgG1_7_log	-0,50	0,77	0,00	0,00
Immuno_IgG2_7_log	-0,44	0,80	-0,07	0,05
Immuno_IgG3_7_log	-0,37	0,71	-0,09	0,07
Immuno_IgG4_7_log	-0,40	0,73	0,06	-0,02
Immuno_IgE_7_log	-0,44	0,77	0,07	-0,08
Immuno_IgA_7_log	0,10	0,20	0,03	-0,21
Immuno_IgM_7_log	-0,01	0,31	0,07	0,00
Immuno_IgG_7_log	-0,50	0,81	-0,03	0,03
Leptin_7_log	0,49	0,20	-0,14	-0,46
Adiponectin_7_log	-0,04	0,20	-0,28	-0,03
Cito_IL27_15_log	-0,10	0,05	0,00	-0,02
Cito_IL1b_15_log	0,41	0,21	-0,36	0,38
Cito_IL2_15_log	0,44	0,33	0,31	0,25
Cito_IL4_15_log	0,03	-0,12	-0,17	0,25
Cito_IL5_15_log	-0,05	0,00	-0,01	0,69
Cito_IL6_15_log	0,42	0,37	-0,43	0,27
Cito_IL10_15_log	0,36	0,38	-0,37	0,15
Cito_IL12_15_log	-0,29	0,16	-0,08	0,51
Cito_IL13_15_log	0,29	0,47	-0,24	0,14
Cito_IL17_15_log	0,20	0,34	-0,27	0,15
Cito_Ifn.g_15_log	0,15	0,37	-0,41	0,60
Cito_GM.CSF_15_log	-0,13	-0,03	0,01	0,14
Cito_TNF.a_15_log	0,32	0,01	-0,29	0,57
Cito_IL9_15_log	-0,11	0,05	-0,05	-0,02
Cito_IL23_15_log	0,37	-0,08	0,51	0,28
Cito_IL18_15_log	-0,10	0,44	-0,20	0,58
Cito_IL21_15_log	-0,11	0,33	0,00	-0,32
Cito_IL22_15_log	-0,22	0,53	0,12	0,11
Immuno_IgG1_15_log	0,81	-0,01	-0,44	-0,09
Immuno_IgG2_15_log	0,70	0,01	-0,41	-0,28
Immuno_IgG3_15_log	0,74	0,09	-0,41	0,09
Immuno_IgG4_15_log	0,68	-0,06	-0,21	-0,22
Immuno_IgE_15_log	0,73	0,20	-0,39	-0,19
Immuno_IgA_15_log	0,16	0,38	0,01	-0,31
Immuno_IgM_15_log	0,00	0,29	-0,04	-0,43
Immuno_IgG_15_log	0,82	-0,02	-0,45	-0,13
Leptin_15_log	-0,01	0,51	-0,01	-0,37
Adiponectin_15_log	0,47	0,04	-0,31	0,05

Tabla 15: MFA para los dos tiempos: Correlación de cada variable con cada una de las dimensiones

A.4. MFA para el incremento entre tiempos

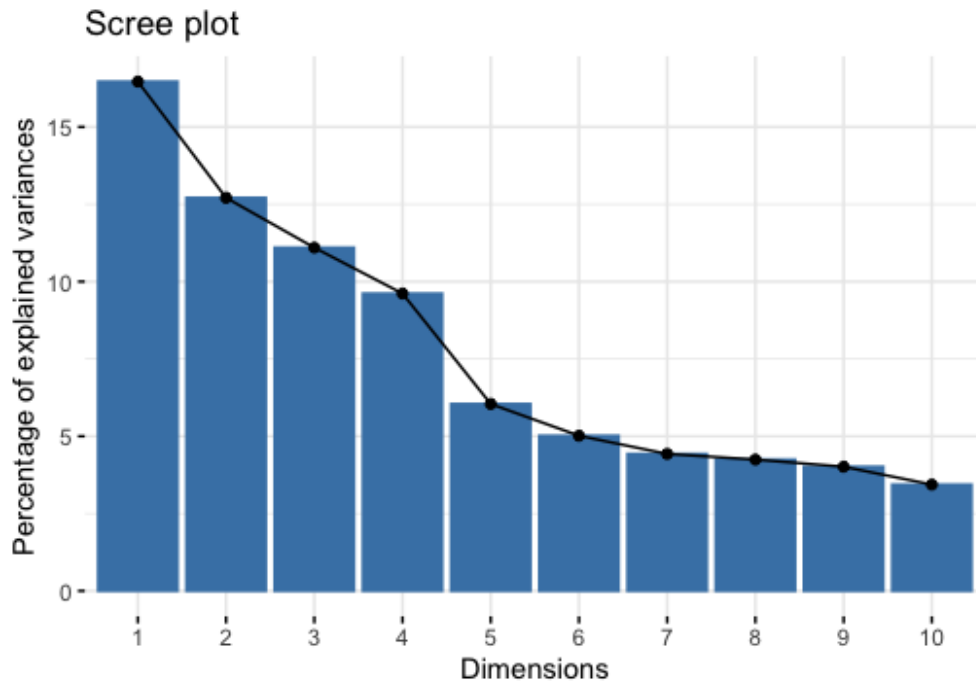


Figura 11: MFA para el incremento entre tiempos: Porcentajes de variabilidad explicados por cada dimensión

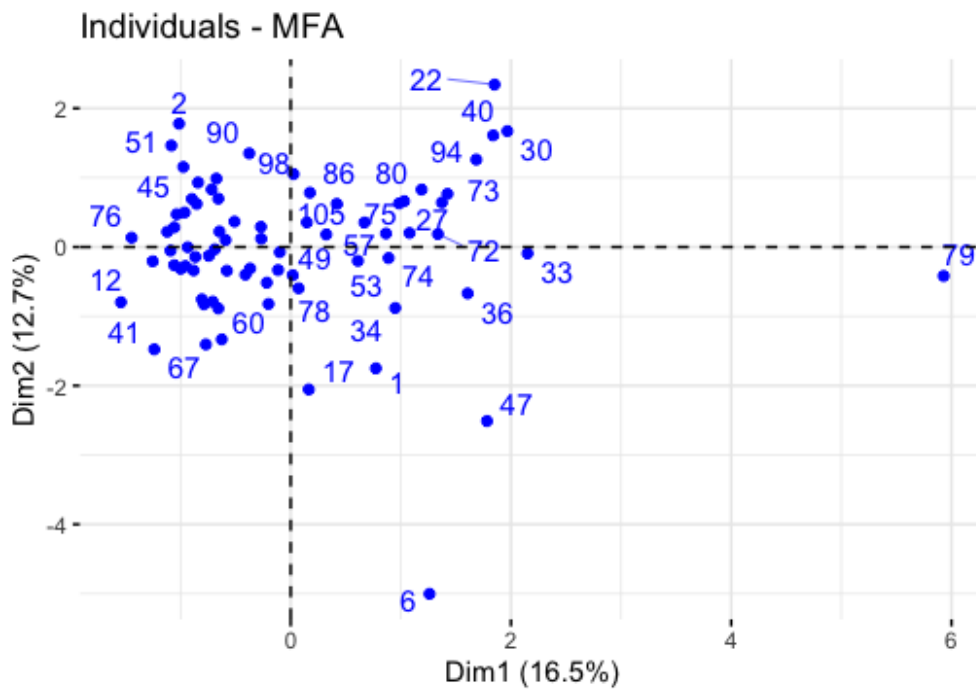


Figura 12: MFA para el incremento entre tiempos: Gráfico de los individuos (1a y 2a dimensiones)

Anexo B. Resultados del Objetivo 2

B.1. Análisis descriptivo inicial

	N	Media	Desv. Est.	Mediana	Mínimo	Máximo	Error Est.
Gestational age	75	39,54	1,24	39,86	36,00	41,57	0,14
Pregnancy weight gain	75	11,83	4,28	12,00	-6,40	20,00	0,49
Pre-gestational BMI	75	22,63	3,34	21,89	17,61	35,76	0,39
Breastfeeding time	75	8,30	3,95	9,00	0,00	12,00	0,46
Exclusive breastfeeding	75	4,00	2,14	5,00	0,00	7,00	0,25
Birth weight	75	3,22	0,46	3,19	2,30	4,64	0,05
BMI Z-score at birth	75	-0,37	0,93	-0,38	-2,47	1,97	0,11
Weight 7D	73	3,21	0,42	3,19	2,40	4,32	0,05
BMI Z-score 7D	73	-0,60	0,88	-0,52	-2,32	1,56	0,10
Weight 15D	71	3,46	0,45	3,40	2,43	4,70	0,05
BMI Z-score 15D	71	-0,58	0,94	-0,63	-2,86	1,32	0,11
Weight 1 M	73	4,07	0,53	3,99	3,12	5,53	0,06
BMI Z-score 1 M	73	-0,66	0,96	-0,69	-3,36	1,68	0,11
Weight 2M	72	5,09	0,65	5,03	4,00	6,70	0,08
Weight 4M	72	6,37	0,72	6,31	4,70	7,80	0,09
Weight 6 MM	73	7,39	0,82	7,30	5,55	9,20	0,10
BMI Z-score 6M	73	-0,38	0,83	-0,44	-2,13	1,47	0,10
Weight 12 M	72	9,33	1,02	9,36	7,06	11,52	0,12
BMI Z-score 12 M	72	0,12	0,93	0,24	-2,21	2,10	0,11
Weight 18 M	75	10,71	1,14	10,75	8,50	13,96	0,13
BMI Z-score 18 M	75	0,30	0,92	0,32	-2,43	2,41	0,11
Weight 24M	32	12,09	1,47	11,56	10,30	16,00	0,26
BMI Z-score 24M	32	0,47	1,04	0,28	-1,46	2,44	0,18

Tabla 16: Estadísticos iniciales (variables numéricas)

	N	Media	Desv. Est.	Mediana	Mínimo	Máximo	Error Est.
Gestational.age_log	75	3,70	0,03	3,71	3,61	3,75	0,00
Pregnancy.weight.gain_log	74	2,52	0,35	2,57	1,10	3,05	0,04
Pre.gestational.BMI_log	75	3,15	0,13	3,13	2,92	3,60	0,02
Breastfeeding.time_log	75	2,10	0,59	2,30	0,00	2,57	0,07
Exclusive.breastfeeding_log	75	1,45	0,66	1,79	0,00	2,08	0,08
Birth.weight_log	75	1,43	0,11	1,43	1,19	1,73	0,01
BMI.Z.score.at.birth_log	58	-0,23	0,73	-0,20	-2,66	1,09	0,10
Weight.7D_log	73	1,43	0,10	1,43	1,22	1,67	0,01
BMI.Z.score.7D_log	48	-0,39	0,90	-0,23	-2,81	0,94	0,13
Weight.15D_log	71	1,49	0,10	1,48	1,23	1,74	0,01
BMI.Z.score.15D_log	51	-0,55	1,14	-0,30	-4,61	0,84	0,16
Weight.1.M_log	73	1,62	0,10	1,61	1,42	1,88	0,01
BMI.Z.score.1.M_log	46	-0,60	1,30	-0,22	-4,61	0,99	0,19
Weight.2M_log	72	1,80	0,11	1,80	1,61	2,04	0,01
Weight.4M_log	72	1,99	0,10	1,99	1,74	2,18	0,01
Weight.6.MM_log	73	2,12	0,10	2,12	1,88	2,32	0,01
BMI.Z.score.6M_log	56	-0,29	0,75	-0,18	-2,41	0,90	0,10
Weight.12.M_log	72	2,33	0,10	2,34	2,09	2,53	0,01
BMI.Z.score.12.M_log	61	0,19	0,60	0,32	-1,20	1,13	0,08
Weight.18.M_log	75	2,46	0,10	2,46	2,25	2,71	0,01
BMI.Z.score.18.M_log	69	0,18	0,76	0,35	-3,22	1,23	0,09
Weight.24M_log	32	2,57	0,11	2,53	2,43	2,83	0,02
BMI.Z.score.24M_log	29	0,30	0,71	0,48	-1,17	1,24	0,13

Tabla 17: Estadísticos después de la transformación (variables numéricas)

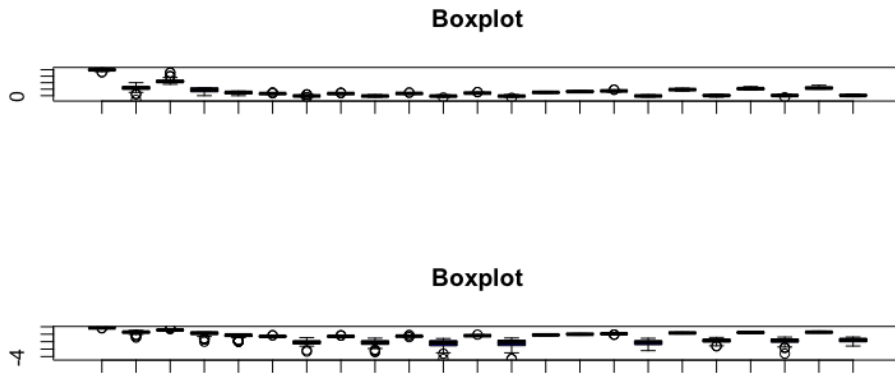
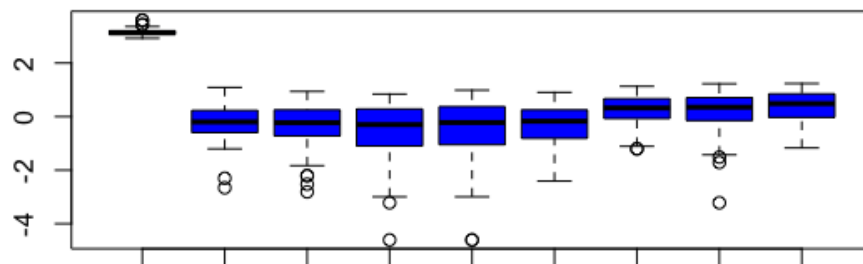


Figura 13: Boxplots múltiples: Comparación entre antes y después de la transformación (variables numéricas)

B.2. Estudio longitudinal: peso, BMI e infecciones

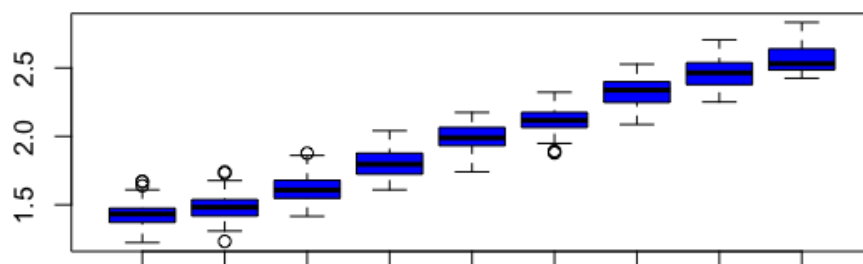
Boxplots múltiples: BMI



Nota: transformado por $\log(x+1)$

Figura 14: Boxplots de la evolución del índice de masa corporal a lo largo de los meses

Boxplots múltiples: Peso



Nota: transformado por $\log(x+1)$

Figura 15: Boxplots del peso a lo largo de los meses

Boxplots múltiples: Infecciones

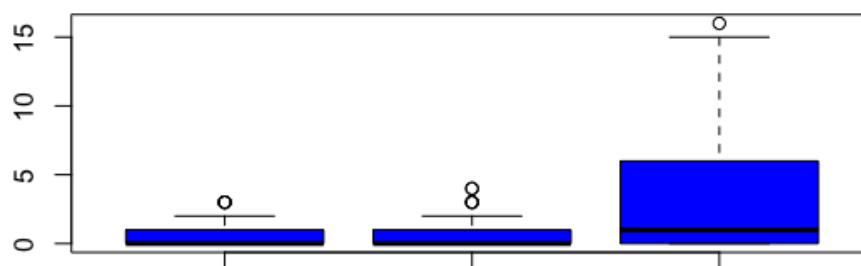


Figura 16: Boxplots de las infecciones a lo largo de los meses

B.3. Comparación entre las dos dietas

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Immuno_IgG1_7	(-0.684,0.129)	0,176	0,230
Immuno_IgG2_7	(-0.556,0.199)	0,346	0,368
Immuno_IgG3_7	(-0.216,0.246)	0,897	0,536
Immuno_IgG4_7	(-0.195,0.247)	0,815	1,000
Immuno_IgE_7	(-0.002,0.001)	0,300	0,564
Immuno_IgA_7	(-0.452,0.228)	0,513	0,281
Immuno_IgM_7	(-0.501,0.586)	0,876	0,965
Immuno_IgG_7	(-0.696,0.186)	0,252	0,313
Immuno_IgG1_15	(-0.445,0.47)	0,955	0,961
Immuno_IgG2_15	(-0.235,0.341)	0,715	0,852
Immuno_IgG3_15	(-0.157,0.354)	0,441	0,466
Immuno_IgG4_15	(-0.22,0.175)	0,824	0,818
Immuno_IgE_15	(-0.001,0.001)	0,629	0,406
Immuno_IgA_15	(-0.154,0.334)	0,464	0,844
Immuno_IgM_15	(-0.87,0.397)	0,458	0,298
Immuno_IgG_15	(-0.394,0.483)	0,840	0,909

Tabla 18: Resultados de las pruebas para la comparación entre las dos dietas (immunoglobulinas)

	p-valor (tobit)	p-valor (Wilcoxon)
Cito_IL27_7	0,643	0,709
Cito_IL1b_7	0,791	0,986
Cito_IL2_7	0,918	0,711
Cito_IL4_7	0,869	0,919
Cito_IL5_7	0,755	0,860
Cito_IL6_7	0,629	0,835
Cito_IL10_7	0,237	0,369
Cito_IL12_7	0,395	0,620
Cito_IL13_7	NA	0,205
Cito_IL17_7	0,514	0,712
Cito_Ifn_g_7	0,776	0,857
Cito_GM_CSF_7	0,230	0,270
Cito_TNF_a_7	0,405	0,509
Cito_IL9_7	0,811	0,859
Cito_IL23_7	0,365	0,306
Cito_IL18_7	0,441	0,908
Cito_IL21_7	0,554	0,521
Cito_IL22_7	0,681	0,908
Cito_IL27_15	NA	0,457
Cito_IL1b_15	0,184	0,257
Cito_IL2_15	0,955	0,923
Cito_IL4_15	0,999	0,457
Cito_IL5_15	0,427	0,311
Cito_IL6_15	0,466	0,077
Cito_IL10_15	0,192	0,202
Cito_IL12_15	0,114	0,147
Cito_IL13_15	0,728	0,739
Cito_IL17_15	0,739	0,816
Cito_Ifn_g_15	0,705	0,551
Cito_GM_CSF_15	NA	0,457
Cito_TNF_a_15	0,789	0,788
Cito_IL9_15	0,411	0,362
Cito_IL23_15	0,357	0,294
Cito_IL18_15	0,546	0,335
Cito_IL21_15	0,236	0,069
Cito_IL22_15	0,084	0,103

Tabla 19: Resultados de las pruebas para la comparación entre las dos dietas (citoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Ratio_L_A_7	(-0.04,0.029)	0,759	0,756
Pro_infl_7	(-0.748,0.553)	0,765	0,814
Anti_infl_7	(-0.972,0.985)	0,990	0,916
Ratio_pro_anti_7	(-1.329,-0.138)	0,017	0,379
Cito_Th1_7	(-0.75,0.523)	0,721	0,443
Cito_Th2_7	(-0.484,0.499)	0,976	0,911
Cito_Rto_Th1_Th2_7	(-0.997,0.271)	0,224	0,269
Th17_7	(-0.973,0.615)	0,653	0,710
Th9_7	(-0.359,0.158)	0,441	0,859
Hematopoyetic_factors_7	(-0.287,0.582)	0,497	0,706
Innate_Immunity_7	(-1.07,0.676)	0,653	0,595
Acquired_immunity_7	(-0.646,0.385)	0,614	0,440
Ratio_Innate_acquired_7	(-1.26,1.878)	0,686	0,655
Immuno_Th1_7	(-0.702,0.172)	0,228	0,259
Immuno_Th2_7	(-0.195,0.247)	0,815	1,000
Immuno_Rto_Th1_Th2_7	(-0.664,0.036)	0,078	0,144
pct_Th1_7	(-0.032,0.005)	0,146	0,144
pct_Th2_7	(-0.042,0.489)	0,098	0,144
pct_Rto_Th1_Th2_7	(-0.664,0.036)	0,078	0,144
Ratio_L_A_15	(-0.076,0.013)	0,165	0,386
Pro_infl_15	(-1.15,0.278)	0,226	0,178
Anti_infl_15	(-1.394,0.392)	0,266	0,077
Ratio_pro_anti_15	(-1.132,1.588)	0,730	0,687
Cito_Th1_15	(-0.735,0.358)	0,493	0,546
Cito_Th2_15	(-0.278,0.365)	0,788	0,959
Cito_Rto_Th1_Th2_15	(-2.334,2.573)	0,899	1,000
Th17_15	(-1.467,0.059)	0,070	0,060
Th9_15	(-0.116,0.106)	0,927	0,362
Hematopoyetic_factors_15	(-0.254,0.315)	0,831	0,553
Innate_Immunity_15	(-1.284,0.359)	0,264	0,178
Acquired_immunity_15	(-0.415,0.191)	0,462	0,264
Ratio_Innate_acquired_15	(-1.465,1.581)	0,937	1,000
Immuno_Th1_15	(-0.39,0.482)	0,834	0,926
Immuno_Th2_15	(-0.248,0.136)	0,563	0,974
Immuno_Rto_Th1_Th2_15	(-0.39,0.446)	0,894	0,683
pct_Th1_15	(-0.019,0.023)	0,861	0,683
pct_Th2_15	(-0.348,0.271)	0,805	0,683
pct_Rto_Th1_Th2_15	(-0.39,0.446)	0,894	0,683

Tabla 20: Resultados de las pruebas para la comparación entre las dos dietas (variables calculadas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Gestational age	(-0.011,0.018)	0,651	0,805
Pregnancy weight gain	(-0.199,0.155)	0,802	0,854
Pre-gestational BMI	(-0.111,0.005)	0,074	0,053
Breastfeeding time	(0.012,0.517)	0,040	0,096
Exclusive breastfeeding	(-0.261,0.378)	0,716	0,309
Birth weight	(-0.044,0.075)	0,600	0,653
BMI Z-score at birth	(-0.561,0.348)	0,636	0,724
Weight 7D	(-0.034,0.069)	0,493	0,541
BMI Z-score 7D	(-0.519,0.618)	0,860	0,548
Weight 15D	(-0.032,0.073)	0,433	0,496
BMI Z-score 15D	(-0.923,0.504)	0,554	0,985
Weight 1 M	(-0.035,0.071)	0,507	0,683
BMI Z-score 1 M	(-0.644,0.839)	0,793	0,856
Weight 2M	(-0.055,0.049)	0,903	0,884
Weight 4M	(-0.028,0.072)	0,377	0,189
Weight 6 MM	(-0.026,0.067)	0,385	0,276
BMI Z-score 6M	(-0.277,0.512)	0,552	0,651
Weight 12 M	(-0.019,0.078)	0,229	0,216
BMI Z-score 12 M	(-0.126,0.486)	0,243	0,226
Weight 18 M	(-0.016,0.077)	0,198	0,214
BMI Z-score 18 M	(-0.304,0.465)	0,675	0,740
Weight 24M	(0,0.167)	0,050	0,065
BMI Z-score 24M	(-0.194,0.955)	0,182	0,033

Tabla 21: Resultados de las pruebas para la comparación entre las dos dietas (variables numéricas)

	p-valor (contraste proporciones)
Infections_0_6M	0,275
Infections_0_12M	0,106
Infections_0_24M	0,284
AB 7d-2M (Infant)	0,590
IAB	0,120
PeAB	0,755
PrAB	0,623
AB7d-2M	0,446

Tabla 22: Resultados de las pruebas para la comparación entre las dos dietas (contadores)

	p-valor (Chi cuadrado)	p-valor (Fisher)
Gender	0,931	0,812
Delivery	0,691	0,618
Breastfeeding	1,000	1,000
Breastfeeding level	0,172	0,195
Primipara	0,567	0,477
Animals	0,271	0,307

Tabla 23: Resultados de las pruebas para la comparación entre las dos dietas (variables categóricas)

B.4. Gráficos de individuos del MFA para ver la relación de los inmunotipos con la dieta y las infecciones

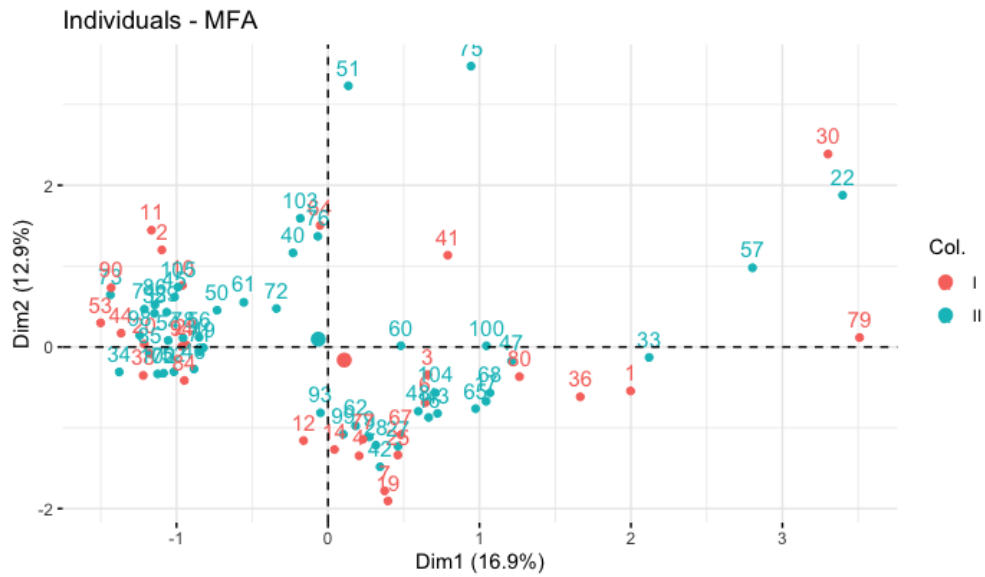


Figura 17: Relación de la dieta con los inmunotipos en el gráfico de individuos

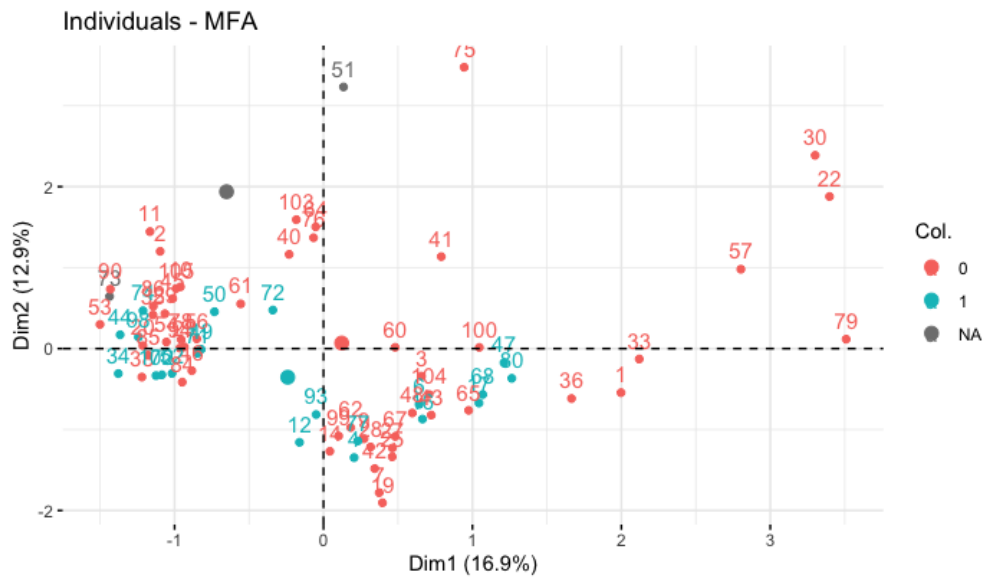


Figura 18: Relación de las infecciones de los 0 a los 6 meses después del parto con los inmunotipos en el gráfico de individuos

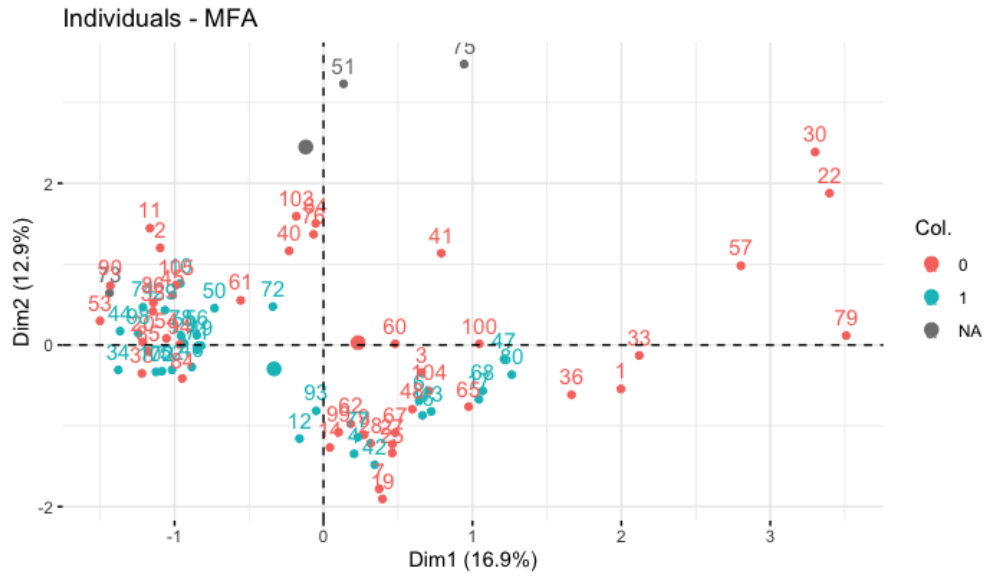


Figura 19: Relación de las infecciones de los 0 a los 12 meses después del parto con los inmunotipos en el gráfico de individuos

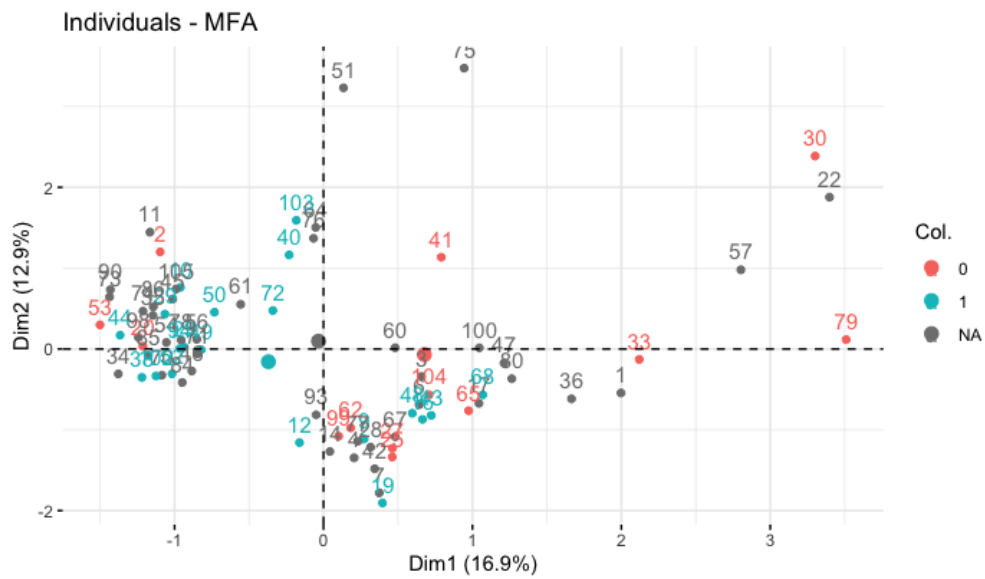


Figura 20: Relación de las infecciones de los 0 a los 24 meses después del parto con los inmunotipos en el gráfico de individuos

B.5. Comparación entre los dos grupos de madres (con outliers)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Immuno_IgG1_7	(1.132,1.575)	2,83E-19	2,65E-12
Immuno_IgG_7	(1.196,1.675)	8,01E-18	2,51E-15
Immuno_IgG_15	(-1.666,-1.138)	2,04E-16	1,17E-14
Immuno_IgG1_15	(-1.706,-1.145)	2,22E-15	1,26E-11
Immuno_IgG2_7	(0.882,1.359)	1,10E-13	7,17E-11
Immuno_IgE_7	(0.004,0.006)	1,72E-12	3,51E-13
Immuno_IgG2_15	(-1.014,-0.589)	1,08E-10	1,25E-08
Immuno_IgG4_7	(0.434,0.75)	1,59E-09	4,19E-09
Immuno_IgG3_7	(0.367,0.677)	3,81E-09	1,44E-09
Immuno_IgE_15	(-0.003,-0.002)	1,66E-08	7,46E-11
Immuno_IgG4_15	(-0.668,-0.344)	8,60E-08	3,10E-08
Immuno_IgG3_15	(-0.801,-0.414)	1,59E-07	1,27E-11
Immuno_IgM_7	(-0.257,0.77)	0,32	0,33
Immuno_IgM_15	(-0.348,0.912)	0,37	0,25
Immuno_IgA_15	(-0.215,0.3)	0,74	0,55
Immuno_IgA_7	(-0.285,0.343)	0,86	0,82

Tabla 24: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (immunoglobulinas)

	p-valor (tobit)	p-valor (Wilcoxon)
Cito_IL18_7	0	0
Cito_IL22_15	0	0
Cito_IL22_7	0	0
Cito_IL10_15	0,01	0
Cito_IL12_15	0,01	0,01
Cito_Ifn_g_7	0,07	0,16
Cito_IL17_7	0,09	0,12
Cito_IL18_15	0,11	0
Cito_IL5_7	0,12	0,11
Cito_IL9_15	0,13	0,02
Cito_IL9_7	0,14	0,09
Cito_TNF_a_15	0,15	0,02
Cito_IL21_7	0,18	0,92
Cito_GM-CSF_7	0,18	0,14
Cito_IL6_15	0,2	0,51
Cito_TNF_a_7	0,24	0,01
Cito_IL6_7	0,24	0,01
Cito_IL21_15	0,36	0,18
Cito_IL1b_15	0,36	0,48
Cito_IL10_7	0,41	0,7
Cito_IL5_15	0,49	0,67
Cito_IL1b_7	0,51	0,99
Cito_Ifn_g_15	0,52	0,65
Cito_IL4_7	0,54	0,69
Cito_IL2_15	0,54	0,52
Cito_IL17_15	0,57	0,49
Cito_IL2_7	0,62	0,78
Cito_IL12_7	0,87	0,17
Cito_IL27_7	0,87	0,95
Cito_IL13_15	0,97	0,98
Cito_IL23_7	1	0,02
Cito_IL23_15	1	0,07
Cito_GM-CSF_15	1	0,35
Cito_IL4_15	1	0,31
Cito_IL13_7	NA	0,31
Cito_IL27_15	NA	0,35

Tabla 25: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (citoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Leptin_7	(-1.25,-0.369)	0,000	0,000
Adiponectin_7	(-0.185,0.579)	0,308	0,270
Leptin_15	(-0.001,1.085)	0,050	0,013
Adiponectin_15	(-0.892,-0.228)	0,001	0,003

Tabla 26: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (adipoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Immuno_Th1_7	(1.171,1.654)	0	0
Immuno_Th1_15	(-1.642,-1.11)	0	0
Immuno_Th2_7	(0.434,0.75)	0	0
Immuno_Th2_15	(-0.643,-0.317)	0	0
Ratio_L_A_15	(0.047,0.135)	0	0
Ratio_L_A_7	(-0.083,-0.013)	0,01	0
Ratio_Innate_acquired_15	(-2.703,-0.346)	0,01	0,02
Pro_infl_7	(-1.309,-0.118)	0,02	0,01
pct_Th1_7	(-0.033,-0.002)	0,03	0,14
Anti_infl_7	(-1.879,-0.057)	0,04	0,02
Hematopoyetic_factors_7	(-0.758,-0.009)	0,05	0,03
pct_Th1_15	(0,0.04)	0,05	0,56
Cito_Rto_Th1_Th2_15	(-3.231,0.217)	0,07	0,11
Innate_Immunity_15	(-1.488,0.067)	0,07	0,07
Th17_7	(-1.388,0.097)	0,09	0,11
Ratio_pro_anti_15	(-1.761,0.25)	0,14	0,92
Hematopoyetic_factors_15	(-0.066,0.458)	0,14	0,41
Cito_Th1_15	(-0.953,0.15)	0,15	0,05
Anti_infl_15	(-1.467,0.24)	0,16	0,08
pct_Th2_7	(-0.09,0.421)	0,2	0,14
Cito_Th1_7	(-0.262,0.94)	0,26	0,02
Innate_Immunity_7	(-1.263,0.385)	0,29	0,27
Pro_infl_15	(-0.339,1.004)	0,33	0,17
Ratio_pro_anti_7	(-0.445,1.257)	0,34	0,96
Th9_7	(-0.391,0.145)	0,36	0,09
Immuno_Rto_Th1_Th2_7	(-0.502,0.197)	0,39	0,14
pct_Rto_Th1_Th2_7	(-0.502,0.197)	0,39	0,14
pct_Th2_15	(-0.434,0.177)	0,4	0,56
Cito_Th2_15	(-0.175,0.412)	0,42	1
Cito_Th2_7	(-0.623,0.296)	0,48	0,62
Acquired_immunity_15	(-0.407,0.21)	0,53	0,01
Th9_15	(-0.08,0.147)	0,55	0,02
Cito_Rto_Th1_Th2_7	(-0.428,0.671)	0,64	0,66
Ratio_Innate_acquired_7	(-1.085,1.66)	0,67	0,72
Immuno_Rto_Th1_Th2_15	(-0.339,0.506)	0,69	0,56
pct_Rto_Th1_Th2_15	(-0.339,0.506)	0,69	0,56
Th17_15	(-0.578,0.85)	0,7	0,68
Acquired_immunity_7	(-0.554,0.44)	0,82	0,51

Tabla 27: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (variables calculadas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Altura 7D	(0.011,0.047)	0	0,01
Weight 7D	(0.009,0.096)	0,02	0,02
Altura 15D	(0.003,0.042)	0,02	0,04
Weight 15D	(0.008,0.099)	0,02	0,03
Altura 1 MM	(0.001,0.042)	0,04	0,05
Peso tras embarazo	(0.001,0.114)	0,05	0,08
BMI Z-score 1 M	(-0.069,1.445)	0,07	0,01
BMI Z-score 15D	(-0.076,1.228)	0,08	0,16
Altura	(-0.005,0.038)	0,13	0,09
Pre-gestational BMI	(-0.014,0.102)	0,14	0,29
BMI tras embarazo	(-0.013,0.087)	0,14	0,14
Weight 1 M	(-0.013,0.083)	0,15	0,13
BMI Z-score at birth	(-0.108,0.667)	0,15	0,23
Birth weight	(-0.015,0.084)	0,17	0,17
Gestational age	(-0.004,0.024)	0,17	0,18
BMI Z-score 7D	(-0.382,0.652)	0,6	0,32

Tabla 28: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (variables numéricas)

	p-valor (contraste proporciones)
Infections_0_6M	0,85
Infections_0_12M	0,23
Infections_0_24M	0,02
AB 7d-2M (Infant)	0,93
IAB	0,61
PeAB	0,68
PrAB	0,21
AB7d-2M	0,06

Tabla 29: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (contadores)

	p-valor (Chi cuadrado)	p-valor (Fisher)
Primipara	0,03	0,02
Breastfeeding level	0,06	0,06
Breastfeeding	0,12	0,1
Animals	0,53	0,56
Gender	0,87	0,82
Delivery	0,99	0,81

Tabla 30: Resultados de las pruebas para la comparación entre los grupos de madres con outliers (variables categóricas)

B.6. Comparación entre los dos grupos de madres (sin outliers)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Immuno_IgG_7	(1.321,1.671)	2,80E-24	3,65E-19
Immuno_IgG1_7	(1.192,1.554)	3,91E-23	2,64E-12
Immuno_IgG2_7	(1.023,1.36)	1,79E-21	3,69E-12
Immuno_IgG_15	(-1.552,-1.159)	6,81E-21	1,46E-16
Immuno_IgG1_15	(-1.576,-1.156)	1,08E-19	6,78E-12
Immuno_IgE_15	(-0.003,-0.002)	2,33E-16	9,73E-12
Immuno_IgE_7	(0.004,0.006)	3,51E-15	1,39E-12
Immuno_IgG3_7	(0.464,0.701)	2,72E-14	7,91E-11
Immuno_IgG2_15	(-0.982,-0.597)	1,39E-11	4,77E-09
Immuno_IgG4_7	(0.423,0.709)	8,63E-10	7,32E-09
Immuno_IgG3_15	(-0.605,-0.347)	4,44E-09	7,93E-11
Immuno_IgG4_15	(-0.668,-0.34)	5,21E-07	5,75E-08
Immuno_IgM_7	(-0.299,0.8)	0,37	0,39
Immuno_IgM_15	(-0.419,0.976)	0,43	0,2
Immuno_IgA_15	(-0.187,0.372)	0,51	0,84
Immuno_IgA_7	(-0.318,0.368)	0,88	0,9

Tabla 31: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (inmunoglobulinas)

	p-valor (tobit)	p-valor (Wilcoxon)
Cito_IL22_15	1,33E-05	9,30E-06
Cito_IL22_7	0	0
Cito_IL18_7	0	0,03
Cito_IL10_15	0,01	0
Cito_IL18_15	0,02	5,03E-05
Cito_IL6_7	0,03	0,02
Cito_TNF_a_7	0,04	0
Cito_IL12_15	0,07078765	0,07353838
Cito_IL13_15	0,07078765	NA
Cito_IL21_7	0,14052368	0,62599369
Cito_IL9_15	0,20031196	0,05837857
Cito_IL17_7	0,22424283	0,28485558
Cito_IL6_15	0,25366303	0,73530924
Cito_IL5_7	0,25772336	0,26538253
Cito_GM-CSF_7	0,29389539	0,23894735
Cito_Ifn_g_15	0,29633577	0,4198717
Cito_TNF_a_15	0,31381967	0,02674385
Cito_IL9_7	0,32031902	0,20217263
Cito_Ifn_g_7	0,33679605	0,60324133
Cito_IL21_15	0,38969076	0,1085043
Cito_IL1b_15	0,63377237	0,93749154
Cito_IL2_15	0,63377237	NA
Cito_IL2_7	0,76304712	0,69200138
Cito_IL1b_7	0,77448481	0,65374976
Cito_IL10_7	0,80075577	0,41951603
Cito_IL12_7	0,82555122	0,39813568
Cito_IL13_7	0,82555122	NA
Cito_IL23_7	0,9975721	0,02058654
Cito_IL17_15	0,99824778	0,11142409
Cito_GM-CSF_15	0,99896176	0,37572106
Cito_IL4_15	0,99914636	0,2877958
Cito_IL5_15	NA	0,11142409
Cito_IL23_15	NA	0,12462481
Cito_IL4_7	NA	0,19946773
Cito_IL27_7	NA	0,37572106
Cito_IL27_15	NA	0,37572106

Tabla 32: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (citoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Leptin_7	(-1.224,-0.332)	0,00	0,00
Adiponectin_7	(-0.088,0.657)	0,13	0,07
Leptin_15	(0.087,1.189)	0,02	0,01
Adiponectin_15	(-0.732,-0.107)	0,01	0,02

Tabla 33: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (adipoquinas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Immuno_Th1_7	(1.297,1.651)	0	0
Immuno_Th1_15	(-1.524,-1.124)	0	0
Immuno_Th2_7	(0.423,0.709)	0	0
Immuno_Th2_15	(-0.638,-0.309)	0	0
Ratio_L_A_15	(0.038,0.126)	0	0
Ratio_L_A_7	(-0.088,-0.014)	0,01	0
pct_Th1_15	(0.003,0.048)	0,03	0,38
Pro_infl_7	(-1.206,-0.059)	0,03	0,03
Ratio_Innate_acquired_15	(-2.883,-0.09)	0,04	0,05
Cito_Th1_7	(0.006,1.123)	0,048	0
pct_Th1_7	(-0.033,0)	0,0531	0,28
Hematopoyetic_factors_15	(-0.007,0.527)	0,06	0,06
Anti_infl_7	(-1.779,0.037)	0,06	0,05
Pro_infl_15	(-0.086,1.271)	0,09	0,04
Th17_7	(-1.336,0.147)	0,11	0,17
Cito_Th2_15	(-0.055,0.464)	0,12	0,38
Cito_Rto_Th1_Th2_15	(-0.055,0.464)	0,12	0,5
Ratio_pro_anti_15	(-2.082,0.31)	0,14	0,87
Hematopoyetic_factors_7	(-0.565,0.084)	0,14	0,08
Innate_Immunity_15	(-1.341,0.206)	0,15	0,11
pct_Th2_15	(-0.52,0.16)	0,29	0,38
Anti_infl_15	(-1.269,0.419)	0,32	0,13
Cito_Th1_15	(-0.82,0.29)	0,34	0,08
pct_Th2_7	(-0.142,0.4)	0,34	0,28
Ratio_pro_anti_7	(-0.503,1.379)	0,35	0,57
Th17_15	(-0.432,1.037)	0,41	0,44
Innate_Immunity_7	(-1.146,0.506)	0,44	0,53
Immuno_Rto_Th1_Th2_15	(-0.337,0.606)	0,57	0,38
pct_Rto_Th1_Th2_15	(-0.337,0.606)	0,57	0,38
Immuno_Rto_Th1_Th2_7	(-0.471,0.269)	0,59	0,28
pct_Rto_Th1_Th2_7	(-0.471,0.269)	0,59	0,28
Ratio_Innate_acquired_7	(-1.156,1.9)	0,62	0,61
Th9_15	(-0.102,0.153)	0,69	0,06
Cito_Rto_Th1_Th2_7	(-0.626,0.774)	0,82	0,81
Cito_Th2_7	(-0.455,0.373)	0,84	0,83
Acquired_immunity_15	(-0.266,0.218)	0,85	0,02
Th9_7	(-0.266,0.22)	0,85	0,2
Acquired_immunity_7	(-0.441,0.414)	0,95	0,34

Tabla 34: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (variables calculadas)

	IC 95 % para la diferencia	p-valor (t Student)	p-valor (Wilcoxon)
Altura 7D	(0.011,0.048)	0	0,01
Weight 7D	(0.004,0.094)	0,03	0,04
Weight 15D	(0.004,0.1)	0,04	0,06
Altura 15D	(0,0.041)	0,06	0,09
BMI Z-score 1 M	(-0.022,1.57)	0,06	0,01
Peso tras embarazo	(-0.004,0.118)	0,07	0,11
Altura 1 MM	(-0.002,0.041)	0,08	0,08
BMI Z-score at birth	(-0.092,0.721)	0,13	0,18
Weight 1 M	(-0.016,0.084)	0,18	0,18
BMI tras embarazo	(-0.018,0.089)	0,19	0,19
Gestational age	(-0.006,0.024)	0,22	0,31
BMI Z-score 15D	(-0.232,0.936)	0,23	0,29
Altura	(-0.009,0.036)	0,24	0,14
Birth weight	(-0.023,0.08)	0,27	0,25
Pre-gestational BMI	(-0.03,0.088)	0,32	0,51
BMI Z-score 7D	(-0.486,0.551)	0,9	0,53

Tabla 35: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (variables numéricas)

	p-valor (contraste proporciones)
Infections_0_6M	0,88
Infections_0_12M	0,48
Infections_0_24M	0,07
AB 7d-2M (Infant)	0,86
IAB	0,76
PeAB	0,91
PrAB	0,59
AB7d-2M	0,07

Tabla 36: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (contadores)

	p-valor (Chi cuadrado)	p-valor (Fisher)
Gender	0,87	0,82
Delivery	0,99	0,81
Breastfeeding	0,12	0,1
Breastfeeding level	0,06	0,07
Primipara	0,03	0,02
Animals	0,53	0,54

Tabla 37: Resultados de las pruebas para la comparación entre los grupos de madres sin outliers (variables categóricas)