



Universidade de Vigo

Trabajo Fin de Máster

Análisis TURF para la extensión de líneas de productos lácteos

Diego García Sánchez

Máster en Técnicas Estadísticas

Curso 2021-2022

Propuesta de Trabajo Fin de Máster

Título en galego: Análise TURF para a extensión de liñas de produtos lácteos
Título en español: Análisis TURF para la extensión de líneas de productos lácteos
English title: TURF analysis for the extension on dairy products lines
Modalidad: Modalidad B
Autor/a: Diego García Sánchez, USC
Director/a: Beatriz Pateiro López, USC
Tutor/a: Maruxa Quiroga García, TasteLab
<p>Breve resumen del trabajo:</p> <p>TasteLab es una start-up de tecnología sensorial basada en la ciencia de los sentidos que se dedica a la prestación de servicios de análisis para la industria del consumo, nace como spin-off de la Universidad de Santiago de Compostela compuesta por un equipo de científicos expertos en análisis sensorial. El objetivo principal del TFM es extender tres líneas de productos lácteos, para posibles clientes. Para ello se recogerá información con consumidores mediante un cuestionario del tipo Best-Worst Scaling, y se realizará un análisis TURF sobre esos datos, buscando optimizar el alcance de los productos ofertados y maximizando los beneficios de venta. En caso de ser necesario, se implementará un procedimiento heurístico para resolver la prueba y reducir el tiempo de procesamiento, más concretamente, del número de niveles de cada atributo y el número de atributos que desee investigar.</p>

Doña Beatriz Pateiro López, profesora Titular de la Universidad de Santiago de Compostela de la USC y doña Maruxa Quiroga García, CEO & Cofounder de TasteLab, informan que el Trabajo Fin de Máster titulado

Análisis TURF para la extensión de líneas de productos lácteos

fue realizado bajo su dirección por don/doña Diego García Sánchez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 02 de febrero de 2022.

La directora:

La tutora:

Doña Beatriz Pateiro López

Doña Maruxa Quiroga García

El autor:

Don Diego García Sánchez

Índice general

Resumen	VII
1. Introducción	1
1.1. TasteLab y el análisis sensorial	1
1.2. Motivación del trabajo	2
1.2.1. Objetivos	2
1.3. Propuesta metodológica	2
2. Marco Teórico	5
2.0.1. Experimentos de Elección Discreta	5
2.1. Análisis TURF	5
2.2. Regresión logística	7
2.2.1. Máxima Verosimilitud	8
2.2.2. Intervalos de confianza basados en el perfil de verosimilitud	9
2.2.3. Contraste de los parámetros	9
2.2.4. Contraste de modelos mediante la deviance	10
3. Desarrollo del proyecto	11
3.1. Asignación de atributos y niveles	11
3.2. Elaboración del cuestionario	12
3.3. Reclutamiento	14
4. Análisis de los datos	15
4.1. Descriptiva	15
4.2. Pretratamiento	18
4.2.1. Regresión logística	18
4.2.2. TURF	18
4.3. Regresión logística	20
4.3.1. Helados	20
4.3.2. Yogures	21
4.3.3. Cremas de queso	23
4.4. TURF	26
4.4.1. Intratabilidad del problema completo	26
4.4.2. Helados	27
4.4.3. Yogures	28
4.4.4. Cremas de queso	29

5. Conclusiones	31
5.1. Interpretación de resultados	31
5.1.1. Perfiles de helado óptimos	31
5.1.2. Perfiles de yogures óptimos	32
5.1.3. Perfiles de cremas de queso óptimos	32
5.2. Críticas al trabajo	32
A. Figuras	35
B. Código	39
B.1. Encuesta	39
B.2. Preprocesado de los datos	48
B.3. Análisis de los datos	55
B.4. Gráficos	60
Bibliografía	61

Resumen

Resumen en español

La extensión de líneas de productos es una tarea apropiada para ser tratada mediante la técnica TURF. A lo largo de este trabajo se aplicará dicha prueba, en conjunción con la regresión logística y un experimento de elección discreta, con el objetivo de seleccionar un conjunto de perfiles de productos lácteos (en concreto helados, yogures y cremas de queso) a través de la técnica TURF para ampliar de la gama de sabores en el mercado de dichos productos. Otros objetivos de este trabajo son la elaboración de una herramienta de recogida de datos y determinar que atributos de los mencionados productos resultan de mayor interés para los consumidores. Se trata de un análisis sensorial llevado a cabo como parte de un programa de prácticas en colaboración con TasteLab. Así, acorde a los intereses de TasteLab se obtiene un conjunto óptimo de sabores (que marcaron tendencia en el mercado) para cada producto, en concreto el conjunto de helados con sabor a fresa, vainilla y frambuesa y mango, alcanzando a un 74,6 % de la muestra; yogures de sabor fresa, arándano, piña y coco (59,3 % de alcance) y cremas de queso sabor ahumado, chocolate y naranja, yogur griego y gin-tonic (alcance del 61 %). Otro hallazgo destacable es la relevancia del etiquetado light en yogures, careciendo de importancia en los demás productos investigados. Se ofrecen también críticas y recomendaciones para la replicación del trabajo.

English abstract

The extension of product lines is an appropriate task to be treated by the TURF technique. Throughout this project, this test will be applied, in conjunction with logistic regression and a discrete choice experiment, with the aim of selecting a set of profiles of dairy products (specifically ice cream, yogurt and cream cheese) through the TURF technique to expand the range of flavors in the market for these products. Other objectives of this paper are the development of a data collection tool and to determine which attributes of the aforementioned products are of greatest interest to consumers. This is a sensory analysis carried out as part of an internship program in collaboration with TasteLab. Thus, according to TasteLab's interest, an optimal set of flavors (which set a trend in the market) is obtained for each product, specifically the set of ice creams with strawberry, vanilla, raspberry and mango flavors, reaching 74.6 % of the sample; strawberry, blueberry, pineapple and coconut flavored yogurts (59.3 % share) and smoked, chocolate and orange cream cheese, Greek yogurt and gin and tonic (61 % share). Another noteworthy finding is the relevance of light labeling in yogurts, lacking importance in the other products investigated. Critiques and recommendations for replication of the work are also offered.

Capítulo 1

Introducción

1.1. TasteLab y el análisis sensorial

Para hablar de análisis sensorial es conveniente definir primero la sensación. La RAE define sensación en su primera acepción como “impresión que percibe un ser vivo cuando uno de sus órganos receptores es estimulado”. Este es, por tanto, el objeto de estudio del análisis sensorial. Esta definición lleva directamente a la necesidad de definir la percepción, la cual podemos definir desde la psicología como una experiencia sensorial consciente, resultado de un proceso perceptual que incluye al estímulo que lo provoca, la señal eléctrica que se transmite por el sistema nervioso del individuo, la experiencia y/o acciones que suceden en el organismo a partir de esta señal eléctrica y el conocimiento e información que las experiencias previas del individuo añaden a este proceso cognitivo. Puede encontrarse una definición más detallada de este proceso, así como una explicación pormenorizada de las diferentes vías perceptivas en Goldstein (2009). Presenta grandes similitudes con la explicación del proceso perceptivo en Meilgaard et al. (2007), donde además puede leerse una guía completa sobre el análisis sensorial. El análisis sensorial se define, por tanto, como el estudio sistemático de la respuesta humana a propiedades psico-químicas, obteniéndose información en cuatro dimensiones de la percepción sensorial: cuantitativa, cualitativa, temporal y hedónica (afectiva), ver Issanchou y Nicklaus (2006). Si bien de forma generalizada para el análisis sensorial es necesario poner a prueba los sentidos de los sujetos experimentales, el hecho de que hablemos de propiedades psico-químicas establece una dimensión que ocurre dentro de la psique del individuo. Por tanto, resulta lógico pensar que siempre que se trate con sensaciones y percepciones ya conocidas previamente, si se las consigue evocar adecuadamente de forma imaginaria, se podría seguir englobando el estudio dentro de esta categoría. Puede leerse más sobre como la capacidad sensorial humana se extiende más allá del presente, permitiendo imaginar y evocar sensaciones, así como sobre la relación entre percepción, sensación e imaginación en Tomlison (2019). De igual manera que para un jugador de baloncesto a punto de lanzar un tiro libre imaginar el movimiento le ayuda a efectuar bien el lanzamiento, investigar las propiedades sensoriales imaginadas de un producto puede llevar a afinar mejor el prototipo sobre el que luego estudiar las propiedades sensoriales percibidas. El interés en el análisis sensorial es en multitud de ocasiones económico. Permite establecer de una manera objetiva el valor de una mercancía y evaluar diferentes aproximaciones para encontrar aquella más rentable. Permite obtener datos fiables y extraer conclusiones en base a las cuales un vendedor pueda tomar decisiones, especialmente en campos como el desarrollo de productos o el control de calidad (Meilgaard, 2006). TasteLab¹ es una start-up de tecnología sensorial basada en la ciencia de los sentidos que se dedica a la prestación de servicios de análisis para la industria del consumo. Nace como spin-off de la Universidad de Santiago de Compostela compuesta por un equipo de científicos expertos en análisis sensorial, el cual se realiza tanto con consumidores como con un panel de catadores entrenados. La información es recogida y analizada de manera automática por

¹<https://tastelab.es/>

su software de diseño propio, *SenseBit*. Ofrecen sus servicios a empresas y pequeños productores que desean conocer la percepción de los consumidores sobre un producto que pretenden implementar. Así, a través de las mencionadas catas, o incluso del uso del software para que el cliente pueda realizar sus estudios por cuenta propia, el cliente puede conocer la percepción que tendrá el consumidor de su producto prototipo, así como si es necesario implementar algún cambio o mejora en el mismo. Se especializan en el sector alimentario, con experiencia en distintos productos de este sector.

1.2. Motivación del trabajo

1.2.1. Objetivos

Este trabajo se enmarca dentro de un proyecto de TasteLab, cuyo objetivo general es determinar los perfiles más atractivos para el público de determinados productos alimenticios, en concreto de helados, yogures y cremas de queso. Puesto que esta nomenclatura de perfiles (con sus atributos y niveles) se va a emplear intensivamente a lo largo del trabajo, conviene explicarla correctamente. Entiéndase como perfil a cada uno de los productos concretos que se plantea sacar al mercado. Se pretende investigar diferentes perfiles, que comparten los atributos o características, aunque no necesariamente todos los valores de los atributos han de ser iguales. Así, cada una de las opciones posibles para un atributo se denominan niveles. Para facilitar la comprensión se expone un ejemplo con calzado deportivo: Se desea establecer cual de los dos perfiles de zapatillas de una marca concreta resultaría más atractivo al público. Los atributos a tener en cuenta son la forma de cierre y sujeción, el tamaño de la suela y si su permeabilidad. Cada una de estas características son los llamados atributos, y cada uno de estos atributos puede presentar varios niveles. De esta forma, el cierre puede ser mediante cordones o velcro, el ancho de la suela es realmente una variable continua, la cual podría ser discretizable en los anchos más frecuentes del mercado (estableciéndose por ejemplo tres niveles, 1,5 cm., 2 cm. y 3 cm.) y aunque la permeabilidad pudiera ser medida de forma continua parece más lógico formularla dicotómicamente (permeable o impermeable). Cada una de las combinaciones concretas de los posibles niveles de los atributos establecería un perfil distinto. Unas zapatillas con velcro, 2cm. de suela e impermeables serían un perfil, mientras que otras con velcro, 3cm de suela e impermeables corresponden a otro perfil distinto.

Se pretende seleccionar un número reducido de perfiles, que maximicen el alcance en el mercado de los productos a tratar. Como objetivos específicos que colaboran a conseguir el objetivo general ya mostrado, además de contar con determinado interés por sí mismos, se destacan:

- Evaluar qué atributos y que niveles de los mismos tienen más relevancia o utilidad a la hora de decantarse por un determinado perfil, especialmente los relativos a atributos de tipo sensorial.
- Seleccionar un conjunto de perfiles para ampliar la línea de productos a través de la prueba TURF, así como a través de otras pruebas, comparando los resultados aportados por todas ellas.
- Elaborar e implementar una herramienta web de recogida de datos adaptada a las necesidades del trabajo, y adaptable con ligeras modificaciones a otros contextos o situaciones similares, mediante la librería shiny.

1.3. Propuesta metodológica

A fin de conocer los diferentes atributos que guían la toma de decisiones del consumidor de determinados productos lácteos se propuso un modelo de selección discreta en el que se muestran (en la herramienta de diseño propio) diferentes niveles de los atributos de interés para esos productos.

Los productos que se investigan son helados, yogures y cremas de queso. El procedimiento que se sigue es similar para los tres, pero los atributos a investigar y los niveles de estos fueron elegidos individualmente para cada uno de ellos.

Con la información recogida se aplican diferentes técnicas estadísticas, a fin de comprobar los resultados que ofrece cada una, tomar a partir de ellas una extensión de líneas de productos y comparar los nuevos perfiles seleccionados por cada una de las pruebas, comparar sus resultados y extraer cualquier otra información que pudiera resultar relevante de cara al objetivo del trabajo.

Las técnicas estadísticas que se aplicarán son:

- TURF (*Total Unduplicated Reach and Frequency*): De especial interés para TasteLab.
- Regresión logística: Habitualmente aplicada en el contexto de experimentos de elección discreta, en los que la variable respuesta es binaria.

Observación: En este TFM se presenta como metodología el modelo de regresión logística si bien, teniendo en cuenta la naturaleza de los datos recogidos, no será adecuado en la práctica ya que no se cumple la hipótesis de independencia (en el cuestionario planteado se recogen varias respuestas para un mismo encuestado). Se recomienda, en su lugar, seguir el planteamiento expuesto en el capítulo 13 de Chapman y McDonell (2019), donde se utiliza un "mixed logit model". Se puede encontrar un ejemplo práctico desarrollado en dicho capítulo.

Capítulo 2

Marco Teórico

2.0.1. Experimentos de Elección Discreta

El proceso de toma de decisiones de un consumidor es uno de los principales objetos de estudio de la investigación en *marketing*. En concreto, es frecuente la investigación respecto a la toma de decisiones entre alternativas con múltiples atributos cada una. Conocer como funciona este proceso, o al menos que resultados produce, permite obtener medidas que faciliten la interpretación de este proceso, facilitando así la modificación de las variables presentes en las alternativas investigadas, a fin de obtener un producto más deseable, o que suscite una mayor intención de compra en el consumidor (Zwerina, 1997).

El experimento de elección discreta que se siguió para este trabajo sigue la línea del expuesto en Chapman y McDonell (2019), en el que se refieren a este tipo de encuestas como una encuesta de conjuntos (*conjoint survey*). Así, en este tipo de encuestas se plantean diferentes alternativas del mismo producto, cada una con unos atributos o características propios, y se solicita que se escoja una concreta. Este esquema se repite un número determinado de veces, que depende de la cantidad de atributos a investigar, de la fiabilidad que se desee conseguir y de los requerimientos cognitivos y atencionales que plantee la propia encuesta, pues puede terminar por ser tediosa para aquel que responde. Existen otros tipos de análisis conjunto que no se englobarían en esta categoría, como aquellos en los que se pide que se valore un producto en un escala tipo Lickert. Ese tipo de encuesta fue descartado, en favor de esta, por considerarse una forma más natural de tomar decisiones.

2.1. Análisis TURF

El análisis TURF (*Total Unduplicated Reach and Frequency*) es una metodología estadística propuesta por Miaoulis et al. (1990). Es, en palabras del autor, un modelo de extensión de líneas de productos, basado en los conceptos de alcance y frecuencia, tomados del ámbito publicitario. Es una técnica frecuentemente empleada en el sector de la alimentación para ampliar la variedad de productos ofertada por una compañía (Lipovetsky 2008).

En esencia es una técnica combinatoria de fuerza bruta (Conklin et al., 1999), que tiene como objetivo tomar un subconjunto B de elementos, tomados del conjunto $A = \{a_1, a_2, \dots, a_n\}$, siendo B el subconjunto de elementos, de un tamaño k determinado, que maximiza el alcance (número de consumidores a los que llega) y la frecuencia (habitualmente medida en número de unidades vendidas por unidad de tiempo). En la literatura (Kreiger y Green, 2000) solemos encontrar mucha más atención a maximizar el alcance. La frecuencia, cuando se recoge el dato, suele ser utilizada para añadir información adicional sobre los posibles subconjuntos B con mayor alcance.

Supóngase un vendedor online que quiere promocionar su tienda en diferentes redes sociales y plataformas en línea. Posee una lista de las diferentes plataformas en las que plantea anunciarse. Esta podría incluir, por ejemplo, redes sociales como Facebook, Twitter, Instagram, Snapchat, Pinterest,

y ciertos periódicos, como El País, La Voz de Galicia, El Mundo, Faro de Vigo, El Correo Gallego y el ABC. En total cuenta con 11 plataformas en las que está planteando anunciarse. Este conjunto de plataformas sería el denominado conjunto $A = \{a_1, a_2, \dots, a_n\}$. Puesto que no desea pagar más de 4 anuncios diferentes, quiere maximizar el alcance de sus anuncios, de tal forma que consiga que la mayor cantidad de público posible vea su anuncio. Por tanto, debe seleccionar un subconjunto (B) de tamaño 4 que maximice el alcance. Así evitará contratar el anuncio en dos plataformas que compartan gran cantidad de público. Para ello selecciona una muestra de su población objetivo (aquella a la que quiere vender su producto), de tamaño n , a la que le pregunta cual de las plataformas digitales presentes en conjunto A utiliza, y con que frecuencia. Posteriormente calcula todas las combinaciones posibles de 4 elementos tomados del conjunto A (11 plataformas): $\binom{11}{4} = 330$. Una vez tiene esos 330 conjuntos de elementos, uno por uno va comprobando cuantos sujetos de la muestra emplean al menos una de las plataformas digitales de cada uno de los subconjuntos. Cuando ha encontrado el subconjunto B con mayor alcance, es decir, aquel en el que más individuos utilizan al menos una de las plataformas, decide anunciarse en esas cuatro plataformas. Cabría esperar que, en este ejemplo concreto, los consumidores de un periódico digital no sean consumidores de tan solo un periódico, sino que lean diferentes fuentes. Así podría resultar redundante anunciarse en dos periódicos y estaría mejor invertido el anuncio en otra plataforma que llegase a un público más particular, como por ejemplo, Pinterest. Este es, en esencia, el funcionamiento del análisis TURF.

Aunque existen en la literatura análisis TURF con datos de diferentes tipos (Miaoulis et al., 1990; Cameron y Little; 2019, Adler et al., 2010), los datos empleados para esta prueba suelen estar recogidos en una matriz con una columna por cada elemento en el conjunto A y n filas, correspondientes con las respuestas de cada individuo, en la cual se indican los elementos escogidos. Estos se codifican con un 1 para aquellos que escoge y con un 0 para aquellos que no. La herramienta empleada para la recogida de datos pueden ser de distintos tipos. Los más habituales son: escalas tipo Likert, en las cuales quien responde evalúa una característica o un constructo con una puntuación generalmente de 1 a 5 o de 1 a 7 (puede leerse en Ankur et al. (2015) una disertación sobre dichas escalas y su utilización en psicometría); pregunta de intención de compra directa, escalas mejor-peor, u otros procedimientos de elección discreta.

La información recogida es sometida a un procedimiento iterativo, que para un tamaño determinado de B comprueba todos los B posibles hasta encontrar aquel en el que al menos un elemento de B haya sido seleccionado por el mayor número de individuos posibles. Al contrario que en nuestro ejemplo, en el cual el anunciante comprobaba el alcance de todos los grupos manualmente, actualmente este procedimiento es asistido computacionalmente.

Puede representarse mediante un diagrama de Venn, donde cada círculo corresponde con los individuos que seleccionan cada elemento de A , y las intersecciones entre estos representan a aquellos individuos que escogen ambos. Sobre este diagrama se busca cubrir la mayor cantidad de superficie tomando un número de elementos de A menor que el total. Se muestra en el diagrama 2.1 ¹ una versión reducida del ejemplo anterior con tan solo 4 periódicos (La Voz de Galicia, El País, Faro de Vigo y el ABC). Si se tuviese que seleccionar un grupo de 3 elementos entre los 4, que maximizase el alcance, parece claro que debería incluir el ABC, pues sus lectores no leen los demás periódicos. El Faro de Vigo y La Voz de Galicia presentan mucho solapamiento, por lo que probablemente sería redundante elegir ambos, y, puesto que la Voz de Galicia y El País apenas presentan solapamiento, cuando este último si lo presenta con El Faro de Vigo, parece lógico pensar que el subconjunto que más área cubre (mayor alcance) es el formado por La Voz de Galicia, El País y el ABC. Nótese que este es un ejemplo ficticio y no tiene por que representar la realidad.

¹Elaboración propia en R con la librería eulerr

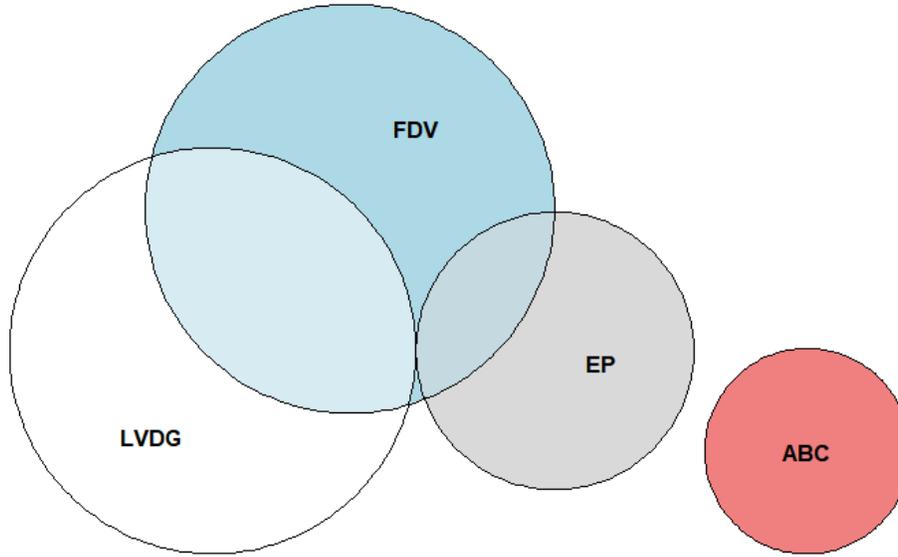


Figura 2.1: Diagrama de Venn representativo del TURF

2.2. Regresión logística

La regresión logística² es un modelo de regresión, derivado del modelo lineal generalizado,

$$Y_i = x_i' \beta + \epsilon_i \text{ para } i \in \{1, \dots, n\}$$

en el cual la variable tomada como respuesta Y es dicotómica, y se codifica como 0 y 1. Esta se pretende explicar mediante una o más variables explicativas X , a veces llamadas covariables. El peso que aporta cada covariable para explicar la variable respuesta está representado por los coeficientes β , y serán estimados por el propio modelo. El error es representado por ϵ_i . Por ser la variable respuesta binaria se considera que sigue una distribución de Bernoulli:

$$E(Y|X = x) = P(Y = 1|X = x) = p(x)$$

en la que definimos el 1 como éxito, 0 como fracaso y p como la probabilidad de éxito.

Sin embargo, expresar la media de Y como una función lineal, condicionada al valor de X $E(Y|X = x) = x_i' \beta$, supuesto que $E(\epsilon) = 0$ resulta incorrecto, ya que incumple las suposiciones básicas del modelo lineal general. Al predecir valores sobre una recta, se obtienen resultados fuera del intervalo $[0, 1]$, en el cual está acotada la variable respuesta, entrando en conflicto con el supuesto de linealidad. El supuesto de homocedasticidad establece que la varianza debe ser constante a lo largo del modelo: $\text{Var}(Y_i|X = x_i) = \sigma$. Sin embargo, la varianza de una variable respuesta binaria $\text{Var}(Y_i|X = x_i) = P(Y_i = 1|X = x_i)[1 - P(Y_i = 1|X = x_i)]$ depende de la probabilidad de éxito, por lo tanto no es igual a lo largo de todos los valores de Y . Por último, como hemos indicado, la distribución de Y es de tipo Bernoulli, incumpliendo el supuesto de que la distribución de la variable respuesta es normal.

A pesar de que la hipótesis de independencia entre los elementos de la muestra si podría ser respetada, el incumplimiento de estos tres supuestos básicos (linealidad, homocedasticidad, normalidad) obliga a plantear el modelo de otra forma.

Para ello, conviene previamente definir la *odds*, como el cociente entre el número de aciertos y el número de fracasos, o equivalentemente, el cociente entre la probabilidad de éxito y la probabilidad de fracaso. $Odds(Y) = P(Y = 1|X = x)/P(Y = 0|X = x)$. Por tanto, estas se mueven en el intervalo

²Aspectos teóricos de esta sección adaptados de Ramil y Saavedra, 2019

$[0, +\infty]$. Puesto que se está buscando una transformación de la variable (Y) que cumpla con los supuestos del modelo lineal generalizado, la Odds(Y) no resulta suficiente (continúa estando en conflicto con el supuesto de linealidad). Pero si se aplica el logaritmo de la Odds, se obtiene una variable respuesta que sí puede tomar valores en el intervalo $[-\infty, +\infty]$, satisfaciendo así el supuesto de linealidad.

Por tanto, podemos definir el modelo logístico como un modelo lineal para el logaritmo de la *odds*. Si se denota

$$\pi(x, \beta) = P(Y = 1 | X = x) \quad (2.1)$$

siendo $\pi(x, \beta)$ la probabilidad de éxito condicionada la variable explicativa y β el valor de un parámetro estimable, se puede observar que la *odds* será

$$\frac{\pi(x, \beta)}{1 - \pi(x, \beta)}$$

El modelo logístico resultante expresa el logaritmo de la odds de la variable respuesta como una función lineal de la variable explicativa, de la forma:

$$\log \frac{\pi(x, \beta)}{1 - \pi(x, \beta)} = x' \beta$$

Si a esta transformación que consiste en aplicar el logaritmo de la *odds* se la llama g , y se la aplica sobre 2.1, se puede definir el modelo logístico como:

$$g(\pi(x, \beta)) = x' \beta \quad (2.2)$$

También es posible invertir la transformación y dejar el modelo expresado en función de la probabilidad de éxito:

$$\pi(x, \beta) = g^{-1}(x' \beta) = \frac{e^{x' \beta}}{1 + e^{x' \beta}} \quad (2.3)$$

2.2.1. Máxima Verosimilitud

Es el método por el que se estiman los coeficientes que acompañan a la/s variable/s explicativa/s. Partamos de una muestra $(x_1, y_1), \dots, (x_n, y_n)$ siendo Y una variable binaria. Como y_i sigue una distribución de Bernoulli $\pi(x_i, \beta)$, se puede expresar la función de máxima verosimilitud como

$$L(\beta) = \prod_{i=1}^n [\pi(x_i, \beta)^{y_i} (1 - \pi(x_i, \beta))^{1-y_i}],$$

su logaritmo como

$$\log L(\beta) = \sum_{i=1}^n [y_i \log \pi(x_i, \beta) + (1 - y_i) \log(1 - \pi(x_i, \beta))]$$

y su derivada parcial respecto a β como

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{\partial \pi(x_i, \beta)}{\partial \beta} \frac{y_i - \pi(x_i, \beta)}{\pi(x_i, \beta)(1 - \pi(x_i, \beta))}.$$

Si ahora se substituye la función genérica $\pi(x, \beta)$ por la expresión a la que ha sido igualada en (2.1), es decir, $\frac{e^{x' \beta}}{1 + e^{x' \beta}}$, se deriva y se iguala a 0 la expresión resultante, se obtiene

$$\frac{\partial \log L(\beta)}{\partial (\beta)} = \sum_{i=1}^n x'_i [y_i - \pi(x_i, \beta)] = 0,$$

la ecuación de máxima verosimilitud, pues igualar a 0 es la condición que se debe cumplir para que $\hat{\beta}$ sea el estimador de máxima verosimilitud.

Sin embargo, como $\pi(x, \beta)$ no es una función lineal de β , es necesario recurrir a un método iterativo para calcular el estimador. En concreto, el software R (R Core Team, 2020) utiliza el método IRLS (*Iterative Reweighted Least Square*). Puede verse una explicación detallada del método, así como una comparación de su desempeño con otros similares en Wolke y Schwetlik (1988).

Una vez estimados los parámetros β del modelo conviene conocer que tipo de inferencia podemos realizar sobre los mismos. En concreto podemos aproximar la distribución del estimador de máxima verosimilitud con una distribución normal, como sigue:

$$\hat{\beta} - \beta_0 \sim N(0, I(\hat{\beta}^{-1})).$$

Aquí se sustituye el parámetro desconocido β por su estimador $\hat{\beta}$, pudiendo considerarse $I(\hat{\beta}^{-1})$ como un estimador de la matriz de covarianzas de $\hat{\beta}$. A partir de esta aproximación se pueden construir regiones de confianza para los estimadores, que toma la forma

$$\{\beta \in \mathbb{R}^p : (\hat{\beta} - \beta_0)^t I(\hat{\beta})(\hat{\beta} - \beta_0) < \chi_{p\alpha}^2\},$$

donde $\chi_{p\alpha}^2$ corresponde con el cuantil $(1 - \alpha)$ de la distribución χ_p^2 . Si a mayores tomamos la matriz $I(\hat{\beta})^{-1}$ y le llamamos $\hat{\sigma}^2(\hat{\beta}_j)$, podemos hallar el error típico de β mediante $\hat{\sigma}(\hat{\beta}_j)$, pues la varianza de cada componente de $\hat{\beta}$ está contenida en ella. Así, con esta estimación del error y la distribución asintótica normal podemos construir intervalos de confianza para cada β_j :

$$(\hat{\beta}_j - z_{\alpha/2} \hat{\sigma}(\hat{\beta}_j), \hat{\beta}_j + z_{\alpha/2} \hat{\sigma}(\hat{\beta}_j))$$

donde $z_{\alpha/2}$ es el cuantil $(1 - \alpha/2)$ de la distribución normal estándar. Sin embargo para muestras pequeñas los estimadores $\hat{\beta}_j$ pueden comportarse de forma asimétrica, por lo que se recomienda construir el intervalo de confianza a través del *profile likelihood*.

2.2.2. Intervalos de confianza basados en el perfil de verosimilitud

El perfil de verosimilitud de un parámetro β_j se define como:

$$PL(\beta_j) = \max_{\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p} L(\beta_1, \dots, \beta_{j-1}, \beta_j, \beta_{j+1}, \dots, \beta_p)$$

Es decir, se maximizan los demás parámetros manteniendo fijo β_j , lo que afecta a la verosimilitud. Los valores de β_j que menos perjudiquen la verosimilitud serán los que formarán parte del intervalo de confianza para β_j . Así se tomará, a nivel de confianza $(1 - \alpha)$, el intervalo formado por:

$$\{\beta_j : 2(PL(\hat{\beta}_j) - PL(\beta_j)) < x_{1,\alpha}^2\}$$

Siendo $x_{1,\alpha}^2$ el cuantil $(1 - \alpha)$ de la distribución ji-cuadrado con un grado de libertad. Por tanto se aplica una aproximación del perfil de verosimilitud por la distribución ji-cuadrado. Este intervalo será asimétrico para muestras pequeñas, y se irá acercando al intervalo obtenido mediante la distribución asintótica a medida que aumente el tamaño muestral.

Debido a que la *odds* y la *oddratio* tienen gran interpretabilidad, puede ser interesante mostrar los intervalos de confianza para este tipo de parámetros. Para ello basta con calcular la exponencial de los parámetros β en los dos extremos del intervalo, como se explicaba en (2.3).

2.2.3. Contraste de los parámetros

Otra inferencia que conviene realizar una vez calculado el modelo es el contraste de los parámetros. Habitualmente se presenta un contraste de hipótesis en el que se toma como hipótesis nula que cierto

parámetro vale 0, y como alternativa que es distinto de 0 ($H_0 : \hat{\beta} = 0$ vs $H_1 : \hat{\beta} \neq 0$). Para este contraste se puede emplear el estadístico de Wald, en el que bajo H_0 :

$$\frac{\hat{\beta}_j}{\hat{\sigma}(\hat{\beta}_j)} \sim N(0, 1)$$

en base al cual se ofrecen habitualmente los niveles críticos para cada coeficiente.

2.2.4. Contraste de modelos mediante la deviance

Un modelo saturado es aquel que proporciona predicciones coincidentes con los valores de la variable respuesta para todos los elementos muestrales, es decir, $\hat{Y}_i = Y_i$ para todo $i \in \{1, \dots, n\}$. No tiene verdadera utilidad estadística, pero es útil como referencia, así como necesario para entender la *deviance*.

Para el contraste entre dos modelos, especialmente logísticos, por la complejidad que conlleva el análisis de los residuos tan habitual en el contexto de la regresión, siendo uno de los cuales una simplificación del otro, se puede emplear la razón de verosimilitud entre dichos modelos. Si designamos como $\hat{\beta}_{H_0}$ la estimación de β bajo la hipótesis nula, y $\hat{\beta}_{H_1}$ como la estimación de β bajo la hipótesis alternativa, el estadístico de contraste toma la forma:

$$-2 \log \frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta}_{H_1})} = -2 \sum_{i=1}^n \left[y_i \log \left(\frac{\pi(x_i, \hat{\beta}_{H_0})}{\pi(x_i, \hat{\beta}_{H_1})} \right) + (1 - y_i) \log \left(\frac{1 - \pi(x_i, \hat{\beta}_{H_0})}{1 - \pi(x_i, \hat{\beta}_{H_1})} \right) \right]$$

La distribución de este estadístico es aproximable por una ji-cuadrado con tantos grados de libertad como parámetros se pierden al pasar de la hipótesis alternativa a la nula.

Por otro lado, la *deviance* es un estadístico de razón de verosimilitud del modelo frente al modelo saturado, es decir:

$$D_{Modelo} = -2 \log \left[\frac{\text{Verosimilitud del modelo}}{\text{Verosimilitud del modelo saturado}} \right]$$

Para la regresión logística, un modelo saturado producirá como ajuste los propios valores y_i . Por tanto, la *deviance* del modelo responde a la fórmula

$$D_{Modelo} = -2 \sum_{i=1}^n \left[y_i \log \left(\frac{\pi(x_i, \hat{\beta}_{Modelo})}{y_i} \right) + (1 - y_i) \log \left(\frac{1 - \pi(x_i, \hat{\beta}_{Modelo})}{1 - y_i} \right) \right]$$

Este estadístico puede ser utilizado como una medida de bondad de ajuste, no de validez del modelo, de la misma manera que se utiliza el R^2 en los modelos lineales. Además se puede obtener el estadístico de razón de verosimilitudes entre dos modelos restando sus *deviances*, de la forma:

$$-2 \log \frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta}_{H_1})} = D_{H_0} - D_{H_1}$$

Por último solo queda obtener un nivel crítico, el cual es calculable en base a la diferencia de *deviances* y de grados de libertad. Así, se resta a la unidad la función de distribución acumulada de la distribución χ^2 para el valor resultante de la diferencia de deviances, con tantos grados de libertad como la diferencia de grados de libertad en los modelos bajo ambas hipótesis. El resultado se compara con la significación deseada para decidir si se acepta o se rechaza el modelo bajo la hipótesis nula.

Capítulo 3

Desarrollo del proyecto

3.1. Asignación de atributos y niveles

La selección de los atributos a investigar para cada uno de los productos de interés, así como la asignación de los respectivos niveles de cada atributo se hizo en base al estudio de tendencias globales en el sector lácteo realizado por el Clúster Alimentario de Galicia (2020). En dicho estudio se analizan las pautas de consumo de productos derivados de la leche a nivel mundial. Se tratan cuestiones como la innovación diferencial en función del país en el que se comercialicen los derivados lácteos, que sabores tienen mejor acogida por el público, la forma de tratar y la acogida que tienen etiquetados y categorías como “listo para consumir”, “sin aditivos” o “bajo en grasa”; o la evolución prevista para este sector alimentario. Puesto que se trata de un estudio detallado, de él se tomaron aquellas categorías de etiquetado más repetidas a lo largo de las innovaciones, las más cercanas al público o las más relevantes para los objetivos de este trabajo.

Aquellos etiquetados y categorías de posicionamiento relativos a la cantidad baja de azúcares o grasas, la no adición de azúcares artificiales o el endulzado mediante diferentes alternativas (agave, dátiles, sirope de arce,...) se recogieron de forma dicotómica bajo el atributo light. De igual modo, con las categorías relativas a la procedencia, tales como el bienestar animal, la no emisión de gases de efecto invernadero, o el etiquetado orgánico se recogieron a través del atributo dicotómico ecológico. Estos dos atributos son comunes a los tres productos, representativos de una tendencia general en el etiquetado de los productos alimenticios. El orden de los niveles (no, sí) no fue elegido de manera aleatoria, sino para, en el análisis posterior, tomar como nivel de referencia el nivel ausente de sendos atributos.

El atributo sabor, también común a los tres productos investigados, y considerado el de principal importancia para esta investigación, presentó más variabilidad entre los distintos productos. Así inicialmente se tomaron aquellos considerados como tendencias, se seleccionaron aquellos que parecían más relevantes, representativos o lógicos, y en ocasiones se modificó la traducción de los mismos realizada por el Clúster Alimentario de Galicia (recordemos que se trata de tendencias a nivel global). Así, la crema de queso sabor vegetales rostizados pasó a ser verduras a la brasa, banano pasó a ser plátano, y otros como el durazno fueron descartados. Se tomó como nivel de referencia el sabor a fresa para helados y yogures, por su ubicuidad. Para las cremas de queso se tomó como referencia el sabor pimentón.

Se añadió también otro atributo para indicar la cantidad a consumir. En el caso del helado, puesto que el consumidor está menos familiarizado con el gramaje, se sustituyó por la forma de presentación. El nivel “cucurucho” de este atributo fue seleccionado como referencia, por ser la forma más habitual de consumo de helado. Para los demás productos se utilizó como nivel de referencia aquel con menor cantidad.

La asignación de atributos y niveles fue la siguiente, en negrita los niveles tomados como referencia en el posterior análisis:

- **Helados**

- Sabor: **Fresa**, Vainilla, Arándanos, Mango, Frambuesa, Plátano, Maracuyá, Coco, Piña, Gin-tonic, Manzana y canela, Calabaza.
- Light: **No**, Sí
- Presentación: **Cucurucho**, Tarrina, Tarrina familiar, Ingrediente de un postre
- Ecológico: **No**, Sí

- **Yogures**

- Sabor: **Fresa**, Vainilla, Arándanos, Mango, Frambuesa, Plátano, Maracuyá, Coco, Piña, Gin-tonic
- Light: **No**, Sí
- Cantidad: **125g.**, 250g., 500g.
- Ecológico: **No**, Sí

- **Cremas de queso**

- Sabor: **Pimentón**, Jalapeño, Ahumado, Crema agria y cebolla, Chocolate y naranja, Verduras a la brasa, Yogurt griego, Plátano, Vino de oporto
- Light: **No**, Sí
- Cantidad: **200g.**, 500g.
- Ecológico: **No**, Sí

Cabe destacar en esta sección que se descartó incluir el atributo precio por dos razones:

- Puesto que se indicaban cantidades diferentes, era necesario indicar una razón precio/cantidad, y no estando el consumidor familiarizado con esta razón en productos previamente envasados (no es así con aquellos productos que se compran a granel como fruta o carne), añadía un nivel más de complejidad a la toma de decisiones. Por tanto, a fin de no confundir al participante del estudio, se omitió la variable precio.
- Incluir el precio en los atributos que se presentaban sobre cada producto implica para el participante enfrentarse a la decisión de si es caro o barato, si lo compraría o no. Así, incluir el precio podría llevar a modificar accidentalmente el objeto de estudio, pasando de ser atractivo percibido a intención de compra.

3.2. Elaboración del cuestionario

Una de las partes fundamentales de este proyecto es la elaboración de la herramienta de recogida de datos. Al carecer Tastelab de una herramienta de recogida de datos acorde a la metodología propuesta, era necesaria la elaboración de dicha herramienta, que permitiese seleccionar conjuntos de perfiles (combinaciones de niveles para cada producto) de forma aleatoria. Se descartó la aplicación de un diseño factorial fraccional ya que este presupone ortogonalidad en los datos, asumiendo por tanto que no existen interacciones entre los distintos niveles. En lugar de ello se optó por conservar todos los perfiles distintos y presentarlos en el cuestionario de forma aleatorizada. Esto condicionó en gran medida la elaboración del cuestionario, al no existir soluciones gratuitas y convenientes. Existen soluciones de pago especializadas en análisis de mercado, tales como Lighthouse Studio (Sawtooth Software, 2021),

sin embargo, se optó por elaborar la herramienta completa, e integrarla en un servidor propio de TasteLab.

Así, esta herramienta no solo facilita la elaboración de este trabajo, sino que aporta valor a TasteLab. Además se creó una herramienta fácilmente adaptable a otras investigaciones y estudios de mercado, de forma que ligeras modificaciones permitan la recogida de datos, en este formato, de cualquier tipo de información relevante a la empresa. Por tanto, podría considerarse un incremento del capital estructural de la empresa, de acuerdo con la definición de Gogan et al. (2015), en la que se define de forma amplia como procedimientos, tecnologías o derechos de propiedad intelectual, o como aquello que queda en la empresa cuando los empleados se van a casa.

Para maximizar la recogida de datos y consecuencia del muestreo por conveniencia que habría que realizar (el cual se esperaba que fuera reducido dadas las condiciones), se estipularon veinte preguntas o comparaciones de perfiles, tomándose estos perfiles de cinco en cinco. De cada conjunto de 5 perfiles, el individuo que responde seleccionaba aquel que le resultaba más atractivo.

La herramienta fue elaborada de forma integral en el lenguaje R, y se utilizaron diferentes librerías para ello, como siguen:

- **shiny**: Una librería especializada en el desarrollo de aplicaciones web interactivas, que permite crear interfaces para el usuario con entradas de datos y establecer una reactividad, de forma que las acciones del usuario puedan modificar la interfaz con la que trata.
- **shinydashboard**: Su función principal es organizar la interfaz de usuario de una aplicación de shiny en pestañas, con un panel lateral que permite navegar entre ellas. Fue usada para establecer cuatro secciones en la encuesta diferenciadas: Información personal, helados, yogures y cremas de queso
- **kableExtra**: Librería utilizada para editar el formato de elementos HTML. Fue utilizada para estilizar las tablas en las que se mostraban el conjunto de alternativas entre los que el individuo debía escoger.
- **rdrop2**: Librería dedicada a permitir la conexión con un directorio de dropbox. Fue utilizada para almacenar las respuestas en dicho directorio.

El cuestionario por tanto se dividió en cuatro apartados. El primero fue utilizado para proporcionar al encuestado las instrucciones para realizar la prueba, así como para recoger la edad, género y frecuencia de consumo de los productos a investigar. Se recogió la frecuencia por ser un objeto de interés en el análisis TURF, mediante una pregunta que ofrecía tres opciones de respuesta: consumo del producto más de una vez por semana, más de una vez por mes y menos de una vez por mes. Se evitó utilizar adjetivos como mucho o poco debido a la subjetividad inherente en ellos. De esta forma se pueden clasificar los consumidores en tres categorías. Consumidores de alta frecuencia, consumidores de frecuencia moderada y consumidores de baja frecuencia.

Los tres apartados subsiguientes fueron iguales en forma. En ellos se presentaba una tabla con cinco perfiles del producto a investigar diferentes, como ya se ha dicho, tomados aleatoriamente. Debajo aparecían cinco botones, y se solicitaba que se seleccionara aquel que se correspondiera con el perfil que resultaba más atractivo. Por último un botón con la palabra continuar escrita guardaba el perfil seleccionado y introducía una nueva tabla con otros 5 perfiles tomados aleatoriamente. Este proceso se repitió 20 veces por cada producto, para cada participante. Puede verse uno de los apartados, a modo de ejemplo de la interfaz de usuario a continuación, figura 3.1, y una muestra de cada una de las secciones del dicho cuestionario en el apéndice, figuras A.1, A.2, A.3 y A.4.

	Opción 1	Opción 2	Opción 3	Opción 4	Opción 5
Sabor	Vainilla	Plátano	Mango	Arándanos	Coco
Light	Si	No	Si	Si	Si
Presentación	Tarrina familiar	Ingrediente de un postre	Tarrina	Ingrediente de un postre	Tarrina
Eco	Si	Si	Si	No	No

La tabla muestra 5 helados distintos, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la forma de presentación y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva

Señala tu opción preferida

Opción 1
 Opción 2
 Opción 3
 Opción 4
 Opción 5

Llevas 0 de 20 contestadas.

Figura 3.1: Interfaz del cuestionario, sección helados

3.3. Reclutamiento

Los candidatos fueron seleccionados mediante un muestreo por conveniencia, a través de distintas fuentes, tales como asociados de TasteLab o las redes sociales del autor (Facebook, Instagram, Whatsapp...). Teniendo en cuenta que en el propio mensaje que se enviaba se agradecía la difusión, y que se consiguió cierto nivel de la misma, el muestreo podría ser considerado una combinación entre el clásico muestreo por conveniencia y un muestreo bola de nieve. Por la naturaleza del muestreo no se fijó un objetivo para el tamaño muestral, sino que se intentó conseguir la participación de la mayor cantidad de individuos posible.

Este tipo de muestreo no probabilístico complica en gran medida la interpretabilidad de los resultados, comprometiendo la validez externa. Es decir, imposibilitando extender las conclusiones más allá de la propia muestra a ninguna población, al contrario que un muestreo probabilístico bien realizado (Borstein et al. 2013). Sin embargo, presenta algunas ventajas respecto a otros tipos de muestreo, tales como la sencillez y la poca demanda de recursos. Esta última fue la razón por la que este tipo de muestreo fue escogido.

Una forma más adecuada de realizar este estudio hubiera sido un muestreo aleatorio sobre la población general de consumidores de productos lácteos. Podría resultar interesante remunerar con una pequeña cantidad monetaria la realización del cuestionario, para asegurar la participación de individuos no relacionados con el proyecto, así como para conseguir un tamaño muestral más grande. Otro punto importante de cara al muestreo, para asegurar la calidad de los resultados y de la interpretación de los mismos, es el tamaño muestral.

Capítulo 4

Análisis de los datos

4.1. Descriptiva

La herramienta diseñada con Shiny recogió un total de 59 respuestas completas (entiéndase por completas a los tres cuestionarios), y un número superior de respuestas parciales. Sin embargo, debido al servidor, algunas de estas respuestas parciales (a la sección helados, o yogures) fueron respondidas por los mismos individuos, al haber tenido ocasionales problemas de conexión que les obligó a volver a entrar. Por tanto, para el análisis sólo se tomarán en cuenta las respuestas completas, evitándose que las respuestas de un mismo individuo aparezcan múltiples veces. Así, la muestra se compone de individuos escogidos por conveniencia que completaron satisfactoriamente los tres cuestionarios en la misma sesión.

La edad media de la muestra es de 36,25 años, con una desviación típica de 16.8091. El rango en el que se mueve va de los 16 a los 78 años. Se trata de 30 varones y 27 mujeres. A pesar de que se ofrecía la opción de seleccionar no decirlo, o género no binario, no fue seleccionada por ningún individuo. A continuación pueden visualizarse estos datos en forma de histograma, figura 4.1, elaborados mediante la librería ggplot2.

Respecto a la frecuencia de consumo de los productos sobre los que se encuestaba, se muestran a continuación unos diagramas de barras que resumen los datos, 4.2a, 4.2b, 4.2c. En ellos se observa que gran cantidad de los individuos tienden a comer helado en temporadas calurosas y yogures más de una vez a la semana. Sin embargo, en las cremas de queso esta tendencia se invierte, la mayoría de individuos la consume muy poco. En todos ellos la opción de consumo más habitual se sitúa a la izquierda, y la menos habitual a la derecha.

Un conteo simple sobre que niveles salieron elegidos mayoritariamente permite hacerse una primera idea de como se distribuye la elección de la muestra, conocer que niveles fueron más populares y que atributos tienen una distribución más desigual. A pesar de que no podamos realizar inferencia a partir de él, ayuda a conocer qué esperar cuando se proceda a ajustar distintos modelos. Para visualizar rápidamente dicho conteo se usará una tabla de frecuencias, las cuales resultan útiles como una primera aproximación, para conocer como se distribuyen los datos. En ellas se puede ver la cantidad de veces que fue escogido cada atributo, así como el porcentaje que supone de entre el total de perfiles escogidos. Se realizó dicho conteo para las tres secciones diferentes del cuestionario, es decir, para los tres productos investigados. Podemos ver la tabla de frecuencias del primero de ellos, los helados, en el cuadro 4.1. En él se observa como los sabores fresa, frambuesa, mango y vainilla son los más escogidos, calabaza tan solo se escoge 19 veces o una ligera tendencia a favor del etiquetado light y ecológico. Las cuatro opciones de forma de presentación es escogieron casi el mismo número de veces, siendo la de tarrina familiar la menos escogida.

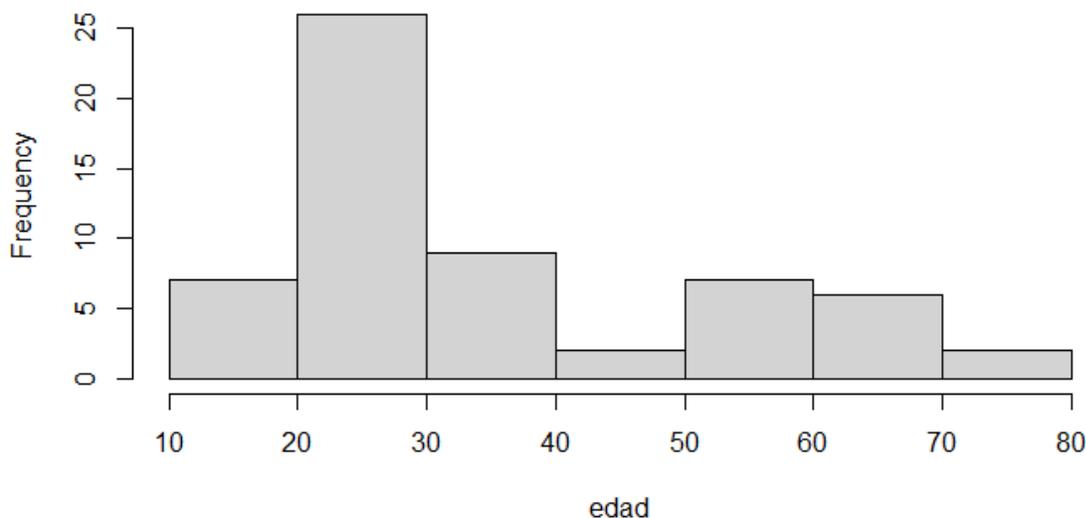


Figura 4.1: Distribución de edades y género de los encuestados

La siguiente tabla de frecuencias a analizar es la relativa a los yogures. Como puede observarse en el cuadro 4.2, a continuación, no aparece ningún sabor que fuera escogido tan pocas veces como el sabor calabaza (hay que tener en cuenta que no era un sabor propuesto en esta sección), sino que se escogieron todos de forma más equilibrada. Se mantiene la tendencia a favor del etiquetado light y ecológico, con más fuerza en ambos casos, pero especialmente para el etiquetado ecológico, el cual fue escogido casi un 59% de las veces. Respecto a las cantidades vemos una relación inversamente proporcional entre la cantidad de producto y la frecuencia con la que este se escoge, indicando una preferencia por empaquetados más pequeños. Cabe recordar que el precio se dejó de lado, y en ningún momento se le indicó a los individuos que una mayor cantidad indicaría una relación precio/cantidad más favorable (como suele ser en el mercado), pero tampoco lo contrario, quedando por tanto a la libre interpretación del encuestado.

Por último, se analiza la tabla de frecuencias relativa a las cremas de queso, la se incluye a continuación, en el cuadro 4.3. Para el atributo sabor de este producto encontramos un claro ganador, el sabor a yogur griego, siendo escogido un 18,64% de las veces. Por la cola destacan el sabor a jalapeño y a plátano, los cuales fueron escogidos un 6,95% y un 6,53% respectivamente. El resto de sabores se escogieron de manera bastante igualitaria. Se mantiene una vez más la tendencia favorable al etiquetado light y ecológico, así como la tendencia a preferir envases con cantidades pequeñas.

Sabores											
Fresa	Arándanos	Calabaza	Coco	Frambuesa	Gin-tonic	Mango	Manzana y canela	Maracuyá	Piña	Plátano	Vainilla
131	103	19	86	166	65	129	69	92	95	66	159
11,10 %	8,73 %	1,61 %	7,29 %	14,07 %	5,51 %	10,93 %	5,85 %	7,80 %	8,05 %	5,59 %	13,47 %

Light		Presentación				Ecológico	
No	Si	Cucurucho	Ingr. postre	Tarrina	T. familiar	No	Si
573	607	301	306	302	271	547	633
48,56 %	51,44 %	25,51 %	25,93 %	25,59 %	22,97 %	46,36 %	53,64 %

Cuadro 4.1: Cuadro de frecuencia: helados

Sabores									
Fresa	Arándanos	Coco	Frambuesa	Gin-tonic	Mango	Maracuyá	Piña	Plátano	Vainilla
182	155	88	169	63	128	100	84	101	110
15,42 %	13,14 %	7,46 %	14,32 %	5,34 %	10,85 %	8,47 %	7,12 %	8,56 %	9,32 %

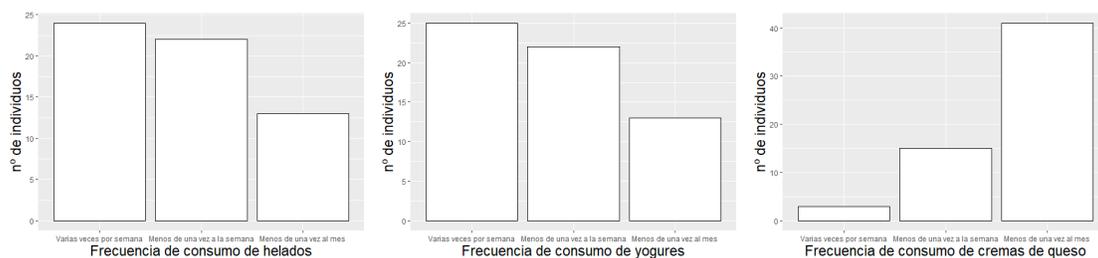
Light		Cantidad			Ecológico	
No	Si	125 g	250 g	500 g	No	Si
558	622	457	386	337	487	693
47,29 %	52,71 %	38,73 %	32,71 %	28,56 %	41,27 %	58,73 %

Cuadro 4.2: Cuadro de frecuencia: yogures

Sabor								
Pimentón	Ahumado	Chocolate y naranja	Crema agria y cebolla	Jalapeño	Plátano	Verduras a la brasa	Vino de Oporto	Yogurt griego
138	145	123	130	82	77	138	127	220
11,69 %	12,29 %	10,42 %	11,02 %	6,95 %	6,53 %	11,69 %	10,76 %	18,64 %

Light		Cantidad		Ecológico	
No	Si	200 g	500 g	No	Si
573	607	631	549	542	638
48,56 %	51,44 %	53,47 %	46,53 %	45,93 %	54,07 %

Cuadro 4.3: Cuadro de frecuencia: cremas de queso



(a) Frecuencias de consumo de helados (b) Frecuencias de consumo de yogures (c) Frecuencias de consumo de cremas de queso

Figura 4.2: Frecuencias de consumo

4.2. Pretratamiento

4.2.1. Regresión logística

Para preparar los datos para la regresión logística, cada una de las respuestas del cuestionario fue almacenada como un objeto de R. Las respuestas de todos los individuos se combinaron después en una matriz de datos (*data.frame*), en el que a cada fila corresponde a una de las opciones que se le ofrecían al individuo que respondía. Las columnas de dicha tabla contienen un identificador del individuo, otro del número de pregunta y otro con el número de alternativa. Seguidos de una columna por cada atributo a investigar, una por cada uno de los datos personales que se encuestaron y por último una en la que se almacena la respuesta. A continuación puede verse la tabla correspondiente a la encuesta acerca de los helados, pero salvo aquellos atributos que cambiaban entre productos, las tablas para yogures y cremas de queso son iguales.

```
> head(long_format.h)
  ID nalt ques.h   Sabor Light   Presentacion Eco edad  genero
1  1  1  1  1 Plátano   No   Tarrina familiar Si  24 Femenino
2  1  2  1  1 Fresa    No   Ingrediente de un postre Si  24 Femenino
3  1  3  1  1 Maracuyá No   Ingrediente de un postre Si  24 Femenino
4  1  4  1  1 Mango     No   Cucurucho No  24 Femenino
5  1  5  1  1 Arándanos No   Ingrediente de un postre No  24 Femenino
6  1  1  2  1 Calabaza No   Cucurucho Si  24 Femenino

      frec.h respuesta
1 Menos de una vez a la semana 0
2 Menos de una vez a la semana 0
3 Menos de una vez a la semana 1
4 Menos de una vez a la semana 0
5 Menos de una vez a la semana 0
6 Menos de una vez a la semana 0
```

4.2.2. TURF

Para el problema TURF, partiendo de la preparación anterior de los datos, se generaron tres matrices de datos, una por cada producto a estudiar. En ellas se incluyeron tantas filas como participantes en la encuesta (59) y tantas columnas como posibles configuraciones de productos existían (192 perfiles de

helado, 120 de yogures y 72 de cremas de queso), a mayores de otras dos columnas, una con la variable ID, que identificaba al respondiente, y otra con la variable peso (*weight*), requerida por el paquete de R que fue usado en el análisis. Esta última se trataba de una fila de unos, pues no existía una razón para asignar pesos diferenciales a distintos individuos (una razón posible hubiera sido asignar mayor peso a los individuos con mayor frecuencia de consumo, o por ejemplo si el objetivo fuese un público de un determinado rango de edad). A aquellos elementos que el cada individuo había seleccionado al menos una vez se les asignó un 1, y a aquellos que no había visto, o que no había seleccionado (en favor de otros) se les asignó un 0. Así, esta es la apariencia que tenía la tabla final (solo se muestran las 20 primeras columnas, pero continúa hasta la columna 194):

```
> head(turfddata.h)
  ID weight p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20
1  1      1  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
2  2      1  0  1  0  0  1  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
3  3      1  0  1  0  0  0  1  0  0  0  0  0  0  0  1  0  0  0  0  0  0
4  4      1  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  1  0  0  1  0
5  5      1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

Como se explicará a continuación, plantear el análisis TURF utilizando esta matriz de datos resultó imposible debido al tamaño del problema y el elevado requerimiento de recursos. Debido a ello, se preparó una versión simplificada de la misma, en la que en lugar de existir una columna para cada uno de los posibles perfiles, existía una columna por cada uno de los sabores. Otra diferencia fundamental es que a aquellos elementos que habían sido visto más de una vez no se les asignó un 1 por haber sido seleccionados al menos una vez, sino que se requirió que hubieran sido seleccionados al menos tantas veces como no seleccionados. Este criterio se aplicó debido a la gran cantidad de veces que le fueron presentados algunos sabores a algunos individuos concretos. Por lo demás el procedimiento fue exactamente el mismo, resultando tres matrices de datos, con 59 filas y 14, 12 y 11 columnas respectivamente para helados, yogures y cremas de queso. La apariencia final fue por tanto la siguiente:

```
> head(turfddata.redux.h)
  ID weight Fresa Vainilla Arándanos Mango Frambuesa Plátano Maracuyá Coco Piña Gin-tonic
1  1      1      0          0          1      0          1      0          1      0      0          0
2  2      1      0          0          0      0          0      0          1      0      0          0
3  3      1      0          1          0      0          0      0          0      0      1          0
4  4      1      0          0          0      0          1      0          0      1      0          0
5  5      1      0          0          0      1          1      0          0      0      0          0
6  6      1      0          0          0      1          0      0          0      0      1          0
  Manzana y canela Calabaza
1          0          0
2          0          0
3          0          0
4          0          0
5          0          0
6          1          0
```

4.3. Regresión logística¹

El modelo de regresión logística fue calculado a través de la función de R base `glm()`, indicándose que se calculase la respuesta en función de los cuatro atributos con los que se trabajaba en cada producto. Posteriormente se aplicó un procedimiento de parada *stepwise*, el cual, tratando de reducir el criterio de información de Akaike (AIC por sus siglas en inglés) va eliminando paso a paso las variables (atributos en este caso) que aumentan dicho criterio. Cuando no puede reducirlo más para, y toma ese modelo como el final. Aplicado este criterio, los resultados se muestran y comentan a continuación.

4.3.1. Helados

Encontramos aquí que el procedimiento *stepwise* elimina la forma de presentación y el etiquetado light, por no considerarlo relevante para el modelo. Procede de igual manera con las variables de información personal y de frecuencia de consumo. En concreto, como veníamos comentando en la sección descriptiva, podría esperarse el bajo poder explicativo de la variable presentación, al haber sido seleccionados los cuatro tipos aproximadamente la misma cantidad de veces. Dentro de las variables que sí que se conservan en el modelo final encontramos que determinados sabores tienen mayor poder predictivo que otros. Es necesario comentar que se investigó la interacción entre variables, infructuosamente, por lo que se presenta un modelo sin interacción. Recordemos que el nivel de referencia fijado fue el helado de fresa en cucurucho, no light y no ecológico. Se muestra a continuación el resumen del modelo calculado:

Call:

```
glm(formula = respuesta ~ Sabor + Eco, family = binomial(link = logit),
     data = long_format.h)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9160	-0.6893	-0.6031	-0.2955	2.5907

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.08006	0.10758	-10.040	< 2e-16	***
SaborArándanos	-0.45974	0.15042	-3.056	0.002241	**
SaborCalabaza	-2.24037	0.25545	-8.770	< 2e-16	***
SaborCoco	-0.57730	0.15697	-3.678	0.000235	***
SaborFrambuesa	0.21739	0.13923	1.561	0.118432	
SaborGin-tonic	-0.74345	0.16941	-4.388	1.14e-05	***
SaborMango	-0.01982	0.14552	-0.136	0.891674	
SaborManzana y canela	-0.89807	0.16518	-5.437	5.43e-08	***
SaborMaracuyá	-0.44722	0.15503	-2.885	0.003918	**
SaborPiña	-0.53053	0.15310	-3.465	0.000530	***
SaborPlátano	-0.85025	0.16792	-5.064	4.12e-07	***
SaborVainilla	0.20923	0.14066	1.487	0.136890	
EcoSi	0.21119	0.06667	3.168	0.001537	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5904.7 on 5899 degrees of freedom

¹ Observación: En este apartado se analizan los resultados del modelo de regresión logística si bien, teniendo en cuenta la naturaleza de los datos recogidos, no es el modelo adecuado al no cumplirse la hipótesis de independencia. Se recomienda, en su lugar, seguir el ejemplo práctico desarrollado en el capítulo 13 de Chapman y McDonell (2019), donde se ajusta un "mixed logit model" haciendo uso de la librería `mlogit`

```
Residual deviance: 5638.1 on 5887 degrees of freedom
AIC: 5664.1
```

```
Number of Fisher Scoring iterations: 5
```

En él destacan un elevado rechazo al sabor calabaza, representado por un coeficiente muy negativo. El elevado error estándar de este sabor respecto a los demás indica que fue escogido muy pocas veces, como se había apreciado en la sección descriptiva. También se aprecian rechazos hacia sabores como plátano, manzana y canela o gin-tonic, entre los más destacados. Aparece una clara tendencia favorable hacia el etiquetado ecológico y hacia determinados sabores como frambuesa, vainilla o mango, incluso presentando los dos primeros un coeficiente positivo. El hecho de que otros coeficientes sean negativos no tiene por que indicar que son sabores rechazados, sino menos escogidos que el sabor a fresa. Se puede observar sin embargo que los sabores con coeficientes más positivos tienden a ser también los menos significativos estadísticamente, probablemente debido a la escasez de datos respecto al gran número de variables del modelo.

Se puede contrastar la verosimilitud del modelo respecto a un modelo nulo, sin predictores, como se muestra a continuación. Vemos que la deviance del modelo más complejo se reduce respecto a la del modelo sin predictores, lo cual indica que el ajuste del más complejo es mejor que el ajuste realizado por el modelo sin predictores.

```
> anova(glm.null.h, glm.step.h, test="Chi")
Analysis of Deviance Table

Model 1: res.h.cod ~ 1
Model 2: res.h.cod ~ Sabor + Eco
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      5899      5904.7
2      5887      5638.1 12    266.65 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para interpretar los coeficientes del modelo conviene deshacer el logaritmo al que ha sido sometido. Así se obtendrá la *odds*, o como de probable es que un elemento sea escogido respecto a que no lo sea, manteniéndose el resto de variables iguales. Así, en este caso, es 1.23 veces más probable que se escoja el helado de fresa ecológico respecto al no ecológico. Pueden observarse las *odds* para todas variable que resultaron significativas, así como su intervalo de confianza al 95% en el ??

4.3.2. Yogures

Para los yogures el procedimiento *step* no elimina ninguna variable salvo las de información personal y la frecuencia de consumo, por lo que todas contribuyen al modelo final. Como ya se comentó en la sección descriptiva de los datos, se percibe una preferencia por las cantidades más pequeñas, suficientemente relevante como para aportar información al modelo. También fue comentado el hecho de que para este producto sí parecía haber una clara tendencia a favor de la etiqueta light, la cual fue seleccionada más veces: como vemos, el *step* la conserva y en el modelo tiene un coeficiente positivo y significativo. También destacan el elevado rechazo hacia el sabor gin-tonic, aunque pueden apreciarse más sabores con coeficientes negativos, así como una tendencia favorable hacia el etiquetado ecológico o a empaquetados más pequeños. Tampoco se halló interacción entre variables para este producto.

Variable	Odd	IC 2,5 %	IC 97,5 %
Intercepto	0,3395736	0,27416372	0,4181154
SaborArándanos	0,6314504	0,46957589	0,8472381
SaborCalabaza	0,1064194	0,06261966	0,1714105
SaborCoco	0,5614138	0,41175384	0,7623110
SaborFrambuesa	1,2428308	0,94654382	1,6342065
SaborGin-tonic	0,4754710	0,33961245	0,6603511
SaborMango	0,980377	0,73690904	1,3041282
SaborManzana y canela	0,4073567	0,29349575	0,5612659
SaborMaracuyá	0,6394029	0,47095073	0,8652630
SaborPiña	0,5882953	0,43500480	0,7931827
SaborPlátano	0,4273060	0,30612336	0,5917457
SaborVainilla	1,2327250	0,93611839	1,6253031
EcoSi	1,2351510	1,08397372	1,4078072

Cuadro 4.4: Helados: odds e intervalo de confianza

Call:

```
glm(formula = respuesta ~ Sabor + Light + Cantidad + Eco, family = binomial(link = logit),
     data = long_format.y)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0183	-0.7103	-0.5859	-0.4567	2.3081

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.92261	0.11048	-8.351	< 2e-16 ***
SaborArándanos	-0.21534	0.12924	-1.666	0.095683 .
SaborCoco	-0.87238	0.14650	-5.955	2.61e-09 ***
SaborFrambuesa	-0.02272	0.12815	-0.177	0.859267
SaborGin-tonic	-1.23883	0.16096	-7.696	1.40e-14 ***
SaborMango	-0.41438	0.13463	-3.078	0.002085 **
SaborMaracuyá	-0.76947	0.14147	-5.439	5.35e-08 ***
SaborPiña	-0.88811	0.14859	-5.977	2.27e-09 ***
SaborPlátano	-0.73620	0.14142	-5.206	1.93e-07 ***
SaborVainilla	-0.56774	0.13900	-4.084	4.42e-05 ***

```

LightSi          0.15293    0.06641    2.303 0.021292 *
Cantidad250 g   -0.26353    0.07933   -3.322 0.000894 ***
Cantidad500 g   -0.42988    0.08158   -5.269 1.37e-07 ***
EcoSi           0.38319    0.06690    5.728 1.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5904.7 on 5899 degrees of freedom
Residual deviance: 5702.8 on 5886 degrees of freedom
AIC: 5730.8

```

Number of Fisher Scoring iterations: 4

Destacan en este modelo la tendencia a favor del etiquetado light y el ecológico, así como el hecho de que ningún sabor sea escogido por encima del sabor a fresa. Especialmente negativos son los coeficientes asociados a los sabores a gin-tonic, piña o coco. Así mismo la gran mayoría son significativos, salvo el sabor a frambuesa. Recordemos que el test de Wald, en el que se basa la significación que sale representada en el resumen del modelo indicaba si se podía asegurar que el coeficiente era distinto de 0.

A continuación se puede observar el contraste de verosimilitud del modelo respecto al modelo nulo, sin predictores. Vemos una reducción de la *deviance* en el modelo más complejo, lo que indica que el ajuste de este mejora al del modelo sin variables predictoras.

```

> anova(glm.null.y, glm.step.y, test = "Chi")
Analysis of Deviance Table

Model 1: res.y.cod ~ 1
Model 2: res.y.cod ~ Sabor + Light + Cantidad + Eco
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      5899      5904.7
2      5886      5702.8 13    201.92 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Por último, se puede observar en el 4.5 que resume todas las variables que entran en el modelo, sus *odds* y el intervalo de confianza para dichas *odds*, a un nivel de confianza del 95%. Destacar de ella la *odd* del helado de frambuesa, muy próxima a 1, indicando que el sabor a frambuesa se escoge prácticamente la misma cantidad de veces que el sabor a fresa. Como se comentaba anteriormente, el coeficiente 0 del sabor a frambuesa está contenido en el intervalo. Destacar también que el yogur de fresa, de 125 gramos, no light y sí ecológico se escogerá 1.47 veces más que su equivalente no ecológico. Este razonamiento es aplicable a todos los elementos del cuadro a continuación.

4.3.3. Cremas de queso

En este caso el procedimiento step solo excluye la variable light, a mayores de las variables de información personal y de frecuencia de consumo, las cuales vemos excluidas en los tres modelos. Se

Variable	<i>Odd</i>	IC 2,5 %	IC 97,5 %
Intercepto	0,3974806	0,3195190	0,4927731
SaborArándanos	0,8062686	0,6254762	1,0383664
SaborCoco	0,4179565	0,3126080	0,5554500
SaborFrambuesa	0,9775337	0,7601937	1,2566067
SaborGin-tonic	0,2897218	0,2100683	0,3951651
SaborMango	0,6607504	0,5068447	0,8594485
SaborMaracuyá	0,4632572	0,3501945	0,6100171
SaborPiña	0,4114306	0,3063838	0,5488729
SaborPlátano	0,4789303	0,3621018	0,6306332
SaborVainilla	0,5668066	0,4307851	0,7431497
LightSi	1,1652450	1,0231179	1,3274140
Cantidad250 g	0,7683322	0,6575247	0,8974296
Cantidad500 g	0,6505881	0,5542012	0,7631191
EcoSi	1,4669502	1,2870608	1,6730464

Cuadro 4.5: Yogures: odds e intervalo de confianza

deduce por tanto que no aportan información relevante para el modelo. Destacan también el elevado rechazo al sabor plátano y al sabor jalapeño, así como la tendencia favorable al etiquetado ecológico, común para los tres productos. No se encontró interacción entre ninguna de las variables incluidas en el modelo.

Call:

```
glm(formula = respuesta ~ Sabor + Cantidad + Eco, family = binomial(link = logit),
    data = long_format.q)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9825	-0.6999	-0.6273	-0.4696	2.1697

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.40944	0.10732	-13.133	< 2e-16 ***
SaborAhumado	0.13692	0.13463	1.017	0.309155

```

SaborChocolate y naranja -0.04683 0.13894 -0.337 0.736084
SaborCrema agria y cebolla -0.08540 0.13674 -0.625 0.532280
SaborJalapeño -0.58610 0.15207 -3.854 0.000116 ***
SaborPlátano -0.69099 0.15442 -4.475 7.65e-06 ***
SaborVerduras a la brasa 0.03710 0.13552 0.274 0.784270
SaborVino de Oporto -0.04892 0.13774 -0.355 0.722495
SaborYogurt griego 0.68789 0.12669 5.430 5.64e-08 ***
Cantidad500 g -0.15354 0.06610 -2.323 0.020196 *
EcoSi 0.24413 0.06613 3.691 0.000223 ***

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 5904.7 on 5899 degrees of freedom
Residual deviance: 5753.3 on 5889 degrees of freedom
AIC: 5775.3

```

Number of Fisher Scoring iterations: 4

Destacan también la tendencia favorable hacia el sabor ahumado, verduras a la brasa y especialmente hacia el yogur griego. En esta ocasión los coeficientes son aparentemente más positivos que en los dos modelos anteriores. La razón a la que se debe esto es que se escogió un perfil de referencia (Pimentón, no light, no ecológico y en envase de 200 gramos) cuyo sabor era menos popular, siendo por tanto escogido menos veces (como se puede comprobar en la sección descriptiva). Por tanto, como los coeficientes se calculan de forma relativa al nivel de referencia, al ser este menos popular, es más probable que otras variables del modelo vayan acompañadas de coeficientes positivos.

Puede verse a continuación, al igual que en las secciones anteriores, dedicadas a los otros dos productos lácteos, como el contraste de la verosimilitud del modelo indica que el modelo calculado mejora el ajuste respecto al modelo sin variables predictoras, pues la *deviance* se ve reducida.

```

> anova( glm.null.q, glm.step.q, test = "Chi")
Analysis of Deviance Table

```

```

Model 1: res.q.cod ~ 1
Model 2: res.q.cod ~ Sabor + Cantidad + Eco
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      5899      5904.7
2      5889      5753.3 10    151.43 < 2.2e-16 ***

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Finalmente, en el 4.6 se puede comprobar la *odds* de cada una de las variables que entran finalmente en el modelo, así como el intervalo de confianza para dichas *odds*. La interpretación de las mismas se realiza en la misma línea que las anteriores. El valor de la *odd* indica que es tantas veces más probable que se escoja esa variable respecto a la que funciona como referencia para ese atributo, manteniéndose el resto de atributos iguales al del nivel de referencia. Así, por ejemplo, la crema de queso sabor a yogur griego, no ecológico y en envase de 200 gramos se escogerá entre 1,55 y 2,55 veces más que la crema de queso sabor pimentón, no ecológica y en envase de 200 gramos, manteniendo el nivel de confianza al 95%.

Variable	Odd	IC 2,5 %	IC 97,5 %
Intercepto	0,2442807	0,1973193	0,3005967
SaborAhumado	1,1467321	0,8807898	1,4935715
SaborChocolate y naranja	0,9542520	0,7263027	1,2526560
SaborCrema agria y cebolla	0,9181444	0,7019745	1,2003183
SaborJalapeño	0,5564921	0,4118129	0,7479683
SaborPlátano	0,5010814	0,3689221	0,6763105
SaborVerduras a la brasa	1,0377954	0,7955719	1,3537936
SaborVino de Oporto	0,9522602	0,7265758	1,2472776
SaborYogurt griego	1,9895116	1,5539631	2,5540628
Cantidad500 g	0,8576668	0,7533351	0,9762135
EcoSi	1,2765113	1,1214706	1,4534345

Cuadro 4.6: Cremas de queso: odds e intervalo de confianza

4.4. TURF

4.4.1. Intratabilidad del problema completo

Inicialmente se trató de realizar la prueba TURF para seleccionar qué grupos de 3, 4, 5 y 6 perfiles completos (p. ej. helado de fresa, en cucurucho, no light y no ecológico) tenían un mayor alcance entre el público encuestado. Para ello se pretendía utilizar la primera de las matrices preparadas para el TURF que se muestran en la Sección 4.2.2. Sin embargo, resultó imposible de realizar debido a que el problema era demasiado grande, y el requerimiento de memoria RAM excedía los 95 Gb. Se intentó fraccionar el problema, y tomar los grupos de 3, 4, 5 y 6 por separado, como funciones distintas. Esto tampoco fue posible, pues continuaba siendo demasiado grande el requerimiento de RAM. Es sencillo ver que se trata de un problema computacionalmente intratable, al menos para un equipo corriente. Si queremos tomar grupos de 3 elementos sobre un conjunto de 192, para después compararlos entre ellos y averiguar cual es el que tiene un mayor alcance, el número combinatorio $\binom{192}{3}$ indica que sería necesario almacenar en memoria un vector de 1161280 elementos, a mayores de la memoria necesaria para realizar las comparaciones entre cada uno de dichos conjuntos de elementos. En el Cuadro 4.7, a continuación, se muestra el número de elementos del vector que tendría que ser almacenado en memoria para cada uno de los tamaños grupales (k) que se deseaban calcular, así como para cada uno de los productos.

Por lo tanto se procedió a utilizar la versión simplificada, en la que en lugar de con perfiles completos

k	Helados (n=192)	Yogures (n=120)	Cremas de queso (n=72)
3	$\binom{192}{3} = 1161280$	$\binom{120}{3} = 280840$	$\binom{72}{3} = 59640$
4	$\binom{192}{4} = 54870480$	$\binom{120}{4} = 8214570$	$\binom{72}{4} = 1028790$
5	$\binom{192}{5} = 2063130048$	$\binom{120}{5} = 190578024$	$\binom{72}{5} = 13991544$
6	$\binom{192}{6} = 64300886496$	$\binom{120}{6} = 3652745460$	$\binom{72}{6} = 156238908$

Cuadro 4.7: Tamaño del problema TURF

solo se trataba con los sabores de cada uno de los productos, tal y como se adelantaba durante el pretratamiento de los datos. En esta ocasión el tamaño del problema era mucho más manejable, y permitió llevar a cabo la prueba sin inconvenientes. Se utilizó para ello la librería `turfR`, indicándose en su función principal (`turf()`) que se debían de tomar grupos del tamaño ya mencionado (3, 4, 5 y 6). No se aplicaron pesos diferenciales a los individuos, pues no existía una justificación para ello. Se muestra a continuación la salida de dicha función para cada uno de los productos. A pesar de que la función devuelve los 924 grupos con mayor alcance, solo se muestran aquí los 3 primeros. La columna `rchX` indica el alcance del grupo en cuestión, expresado en forma de probabilidad, y la columna `frqX` la frecuencia. Las siguientes columnas corresponden a cada uno de los sabores, que, a pesar de venir indicados como números, mantienen el orden con el que fueron codificados. De todos modos se interpretarán las salidas para facilitar la comprensión del lector. Puede apreciarse también como el tiempo requerido para la computación es, ahora si, reducido. Además, a medida que reducimos el tamaño del conjunto del que vamos a tomar los subconjuntos, el tiempo de procesamiento disminuye (es decir, es más alto para los helados ($n = 12$), y más bajo para las cremas de queso ($n = 9$)).

4.4.2. Helados

```
> turf(turfdata.redux.h, 12, 3:6)
3 of 12: 0.563071 sec
4 of 12: 0.5186219 sec
5 of 12: 0.621335 sec
6 of 12: 0.4607742 sec
total time elapsed: 2.239579 sec
$turf
$turf[[1]]
  combo      rchX      frqX 1 2 3 4 5 6 7 8 9 10 11 12
1      3 0.6610169 0.8644068 1 1 0 0 1 0 0 0 0 0 0 0
2     65 0.6440678 0.8135593 0 1 0 1 1 0 0 0 0 0 0 0
3     75 0.6271186 0.7118644 0 1 0 0 1 0 0 1 0 0 0 0

$turf[[2]]
  combo      rchX      frqX 1 2 3 4 5 6 7 8 9 10 11 12
1     10 0.7457627 1.0847458 1 1 0 1 1 0 0 0 0 0 0 0
2     19 0.7288136 1.0169492 1 1 0 0 1 0 1 0 0 0 0 0
```

```
3      23 0.7288136 0.9661017 1 1 0 0 1 0 0 0 0 0 1 0
```

```
$turf[[3]]
```

	combo	rchX	frqX	1	2	3	4	5	6	7	8	9	10	11	12
1	39	0.7966102	1.2033898	1	1	0	1	1	0	0	1	0	0	0	0
2	42	0.7966102	1.1864407	1	1	0	1	1	0	0	0	0	0	1	0
3	74	0.7966102	1.1186441	1	1	0	0	1	0	1	0	0	0	1	0

```
$turf[[4]]
```

	combo	rchX	frqX	1	2	3	4	5	6	7	8	9	10	11	12
1	98	0.8474576	1.3050847	1	1	0	1	1	0	0	1	0	0	1	0
2	94	0.8305085	1.3389831	1	1	0	1	1	0	1	0	0	0	1	0
3	96	0.8305085	1.3220339	1	1	0	1	1	0	0	1	1	0	0	0

Aquí vemos como el grupo de tamaño 3 ($k = 3$) con mayor alcance, llegando a un 66,1% de la muestra es el formado por el helado sabor a fresa, sabor vainilla y sabor frambuesa, seguido por el grupo formado por los helados sabor vainilla, mango y frambuesa (66,4% de alcance) y en tercer lugar el formado por los helados sabor vainilla, frambuesa y piña (62,7

Para los grupos de tamaño cuatro ($k = 4$), el de mayor alcance en la muestra (74,6%), es el formado por los helados sabor fresa, vainilla, mango y frambuesa. El siguiente en alcance (72,9%) es el formado por los sabores de fresa, vainilla, frambuesa y coco. Con el mismo alcance aparece también el grupo de sabores de fresa, vainilla, frambuesa y manzana y canela.

En los grupos de tamaño 5 ($k = 5$), los tres primeros grupos presentan el mismo alcance (79,7%) son los formados por: a) fresa, vainilla, mango, frambuesa y piña; b) fresa, vainilla, mango, frambuesa y manzana y canela; y c) fresa, vainilla, frambuesa, coco y manzana y canela. Destaca que los tres primeros grupos presentan exactamente el mismo alcance, probablemente debido al reducido tamaño muestral de este estudio. De todos modos, el posicionamiento es correcto ya que la frecuencia del que aparece en primera posición es mayor. También cabe mencionar que el cuarto grupo, el cual no se muestra por simplificar los resultados mostrados, presentaba un alcance ligeramente inferior.

Por último, para los grupos de 6 sabores de helado ($k = 6$), el grupo con mayor alcance (84,7%) es el formado por los sabores a fresa, vainilla, mango, frambuesa, piña y manzana y canela. En adelante, el alcance está empatado en el 83,1% hasta el grupo 12. Se posicionan por tanto en función de la frecuencia. Así, el siguiente grupo es el formado por los helados sabor a fresa, vainilla, mango, frambuesa, coco y manzana y canela. En tercer lugar se encuentra el grupo de sabores a fresa, vainilla, mango, frambuesa, piña y gin-tonic.

4.4.3. Yogures

```
> turf(turfddata.redux.y, 10, 3:6)
3 of 10: 0.2922099 sec
4 of 10: 0.2952132 sec
5 of 10: 0.2852249 sec
6 of 10: 0.3879631 sec
total time elapsed: 1.314483 sec
$turf
$turf[[1]]
      combo      rchX      frqX 1 2 3 4 5 6 7 8 9 10
1      13 0.5254237 0.5423729 1 0 1 0 0 0 0 1 0 0
2      10 0.4915254 0.6101695 1 0 1 0 1 0 0 0 0 0
```

```
3      9 0.4576271 0.5254237 1 0 1 1 0 0 0 0 0 0
```

```
$turf[[2]]
```

```
      combo      rchX      frqX 1 2 3 4 5 6 7 8 9 10
1      37 0.5932203 0.7118644 1 0 1 0 1 0 0 1 0 0
2      32 0.5593220 0.6271186 1 0 1 1 0 0 0 1 0 0
3      48 0.5593220 0.6101695 1 0 1 0 0 0 0 1 0 1
```

```
$turf[[3]]
```

```
      combo      rchX      frqX 1 2 3 4 5 6 7 8 9 10
1      80 0.6271186 0.7796610 1 0 1 0 1 0 0 1 0 1
2      79 0.6271186 0.7627119 1 0 1 0 1 0 0 1 1 0
3      59 0.6101695 0.7966102 1 0 1 1 1 0 0 1 0 0
```

```
$turf[[4]]
```

```
      combo      rchX      frqX 1 2 3 4 5 6 7 8 9 10
1     100 0.6610169 0.8305085 1 0 1 0 1 0 0 1 1 1
2      79 0.6440678 0.8644068 1 0 1 1 1 0 0 1 0 1
3      24 0.6440678 0.8474576 1 1 1 0 1 0 0 1 0 1
```

Para los grupos de tamaño 3 ($k = 3$) de sabores de yogur, el grupo con mayor alcance (52,5%) es el formado por los yogures sabor fresa, arándanos y coco. El siguiente, con un alcance del 49,2% es el formado por los sabores a fresa, arándanos y frambuesa. En tercer lugar de acuerdo al alcance (45,8%) es el que abarca los yogures sabor a fresa, arándanos y mango.

En los grupos de tamaño 4 ($k = 4$), el de mayor alcance (59,3%) es el formado por los sabores a fresa, arándano, piña y coco. Los dos siguientes presentan el mismo alcance (55,9%). El de mayor frecuencia es el que abarca los sabores a fresa, arándano, mango y coco. El de menor frecuencia, fresa, arándano, coco y gin-tonic.

Para los grupos de tamaño 5 ($k = 5$), aparece en primera posición con un alcance del 62,7% aquel que incluye los sabores a fresa, arándano, frambuesa, coco y gin-tonic. Con el mismo alcance pero una frecuencia ligeramente más baja encontramos el grupo formado por los sabores a fresa, arándano, frambuesa, coco y piña. Con un alcance del 61% está el grupo que incluye los sabores a fresa, arándano, mango, frambuesa y coco.

Por último, entre los posibles grupos de tamaño 6 ($k = 6$), el de mayor alcance (66,1%) es el formado por los yogures sabor a fresa, arándano, frambuesa, coco, piña y gin-tonic. Los dos siguientes tienen alcances iguales (64,4%). Por frecuencia se sitúa primero el que comprende yogures sabor a fresa, arándano, mango, frambuesa, coco y gin-tonic. Después de él se encuentra el grupo formado por fresa, vainilla, arándano, frambuesa, coco y gin-tonic.

4.4.4. Cremas de queso

```
> turf(turfddata.redux.q, 9, 3:6)
3 of 9: 0.138644 sec
4 of 9: 0.2323811 sec
5 of 9: 0.16856 sec
6 of 9: 0.1525888 sec
total time elapsed: 0.7210929 sec
$turf
$turf[[1]]
```

	combo	rchX	frqX	1	2	3	4	5	6	7	8	9
1	48	0.5423729	0.6101695	0	1	0	0	0	0	1	0	1
2	41	0.5254237	0.6271186	0	1	0	0	1	0	1	0	0
3	73	0.5254237	0.6101695	0	0	0	1	0	0	1	0	1

\$turf[[2]]

	combo	rchX	frqX	1	2	3	4	5	6	7	8	9
1	86	0.6101695	0.7796610	0	1	0	0	1	0	1	0	1
2	89	0.6101695	0.7118644	0	1	0	0	0	1	1	0	1
3	106	0.5932203	0.7627119	0	0	1	0	1	0	1	0	1

\$turf[[3]]

	combo	rchX	frqX	1	2	3	4	5	6	7	8	9
1	102	0.6779661	0.8813559	0	1	0	0	1	1	1	0	1
2	85	0.6610169	0.8983051	0	1	1	0	1	0	1	0	1
3	98	0.6610169	0.8474576	0	1	0	1	0	1	1	0	1

\$turf[[4]]

	combo	rchX	frqX	1	2	3	4	5	6	7	8	9
1	32	0.7288136	0.9322034	1	1	0	0	1	1	1	0	1
2	73	0.7118644	1.0169492	0	1	0	1	1	1	1	0	1
3	68	0.7118644	1.0000000	0	1	1	0	1	1	1	0	1

En esta sección se repasan los resultados de la prueba TURF para los sabores de cremas de queso. En primer lugar se revisan los grupos de tamaño 3 ($k = 3$). El de mayor alcance (54,2%) es el que incluye los sabores ahumado, yogur griego y vino de Oporto. Seguido por el grupo formado por sabores ahumado, chocolate y naranja y yogur griego, con un 52,5% de alcance. Por último, con el mismo alcance, aquel que incluye sabor pimentón, yogur griego y vino de Oporto.

En los grupos de tamaño 4 ($k = 4$) aparece en primera posición, con un alcance del 61% la combinación de sabores ahumado, chocolate y naranja, yogur griego y gin-tonic. Con exactamente el mismo alcance pero menor frecuencia, se encuentra la combinación de sabores ahumado, verduras a la brasa, yogur griego y vino de Oporto. Con un alcance del 59,3% vemos el grupo formado por los sabores a crema agria y cebolla, chocolate y naranja, yogur griego y vino de Oporto.

Para los grupos de tamaño 5 ($k = 5$), la combinación con mayor alcance (67,8%) es la que incluye sabores ahumado, chocolate y naranja, verduras a la brasa, yogur griego y vino de Oporto. El segundo grupo por alcance (66,1%) y frecuencia es el que abarca los sabores ahumado, crema agria y cebolla, chocolate y naranja, yogur griego y vino de Oporto. Con un alcance igual, el grupo que comprende los sabores ahumado, pimentón, verduras a la brasa, yogur griego y vino de Oporto.

Finalmente, se revisarán los grupos de tamaño 6 ($k = 6$). El de mayor alcance (72,9%) es grupo de sabores a jalapeño, ahumado, chocolate y naranja, verduras a la brasa, yogur griego, y vino de Oporto. El siguiente por alcance (71,2%) es el que abarca los sabores ahumado, pimentón, chocolate y naranja, verduras a la brasa, yogur griego y vino de Oporto. Con el mismo alcance y una frecuencia ligeramente inferior aparece la combinación de sabores ahumado, crema agria y cebolla, chocolate y naranja, verduras a la brasa, yogur griego y vino de Oporto.

En general destaca que el sabor a plátano no aparece en ninguno de los doce grupos, el sabor a jalapeño tan sólo una vez, el sabor a yogur griego en todos los grupos, y el sabor a vino de oporto en todos salvo en el segundo grupo de tamaño 3.

Capítulo 5

Conclusiones

5.1. Interpretación de resultados

A un nivel general de interpretación surgen un par de puntos cuanto menos curiosos, que podrían incluso sugerir investigaciones posteriores. El primero es que el atributo light solo resultó relevante para los yogures, siendo descartado tanto para las cremas de queso como para los helados. Aunque no se puede asegurar, una hipótesis razonable es que tanto helados como cremas de queso suponen para el consumidor un alimento menos esencial, más singular, y por la propia naturaleza del mismo asumen que su contenido en azúcares o grasas debe ser elevado para que no se pierda la calidad sensitiva de este. Sin embargo, parece más probable que el consumidor esté más familiarizado con yogures etiquetados como light y satisfecho con ellos, o bien que por ser un producto de uso más recurrente prefiera reducir el contenido calórico de la que supone una cena ligera. El segundo es que llama la atención que pese a que se contrastaron modelos logísticos con interacción, la interacción entre variables no resultara significativa en ningún momento. Se contemplaba la posibilidad de interacción, por ejemplo, entre determinados sabores de helado y la forma de presentación de los mismos. Así, se esperaba que un helado de *gin-tonic* en tarrina tamaño familiar resultara poco atractivo, pues al ser un sabor novedoso y al que el consumidor probablemente no esté acostumbrado, una gran cantidad del mismo podría suponer que envejeciera en el congelador. Sin embargo, por esta misma razón parecía el ingrediente ideal para un postre. Es posible que la falta de interacción significativa se deba al reducido tamaño muestral.

A un nivel más concreto, y puesto que el objetivo era ampliar la gama de artículos lácteos a ofertar por un cliente sin haberse definido previamente una gama de artículos ya ofertados, para cada uno de los productos se examinará el grupo de sabores seleccionados por la prueba TURF. Posteriormente se aplicará sobre ellos los resultados del modelo logístico, configurándose así el producto con los cuatro atributos óptimos de acuerdo a los resultados. Para aquellos productos para los cuales algún atributo no ha llegado a entrar en el modelo, por no resultar significativo, se indicará y se dejarán a decisión del cliente.

5.1.1. Perfiles de helado óptimos

Como ya se ha podido leer en la sección 4.3.1, no parece relevante para los helados ni la forma de presentación del mismo ni el etiquetado light, al contrario que para el atributo ecológico, cuyo nivel positivo resultaba favorable. Por lo tanto, cualesquiera que sean los grupos de sabores escogidos por el TURF, deberían ser etiquetados como ecológicos, y tanto la forma de presentación como el etiquetado light quedar sujetos a las necesidades o conveniencias de la producción. En el cuadro 5.1 se puede comprobar cuales serían los grupos óptimos para cada uno de los tamaños tomados por la prueba TURF, así como el alcance sobre la muestra que tendría cada uno de ellos.

Tamaño del grupo	Sabor	Otros atributos	Alcance
k = 3	Fresa, vainilla y frambuesa	Producción ecológica	66,1 %
k = 4	Fresa, vainilla, mango y frambuesa	Producción ecológica	74,6 %
k = 5	Fresa, vainilla, mango, frambuesa y piña	Producción ecológica	79,7 %
k = 6	Fresa, vainilla, mango, frambuesa, piña y manzana y canela	Producción ecológica	84,7 %

Cuadro 5.1: Combinaciones resultantes para helados

5.1.2. Perfiles de yogures óptimos

Tal y como se comentaba en la sección 4.3.2, para los yogures, todos los atributos investigados fueron significativos y se incluyeron en el modelo final. Así, a mayores de escogerse qué sabores deberían ser incluidos entre los ofertados para ampliar la gama de yogures a comercializar de acuerdo a los resultados del análisis TURF, atendiendo a los resultados del modelo logístico sería recomendable que todos ellos fueran etiquetados como light y ecológico, y envasados en dosis individuales de 125 gramos. En el cuadro 5.2 se muestra como quedarían los distintos grupos atendiendo al tamaño de la gama de productos a ofertar, así como el alcance que estos tendrían en la muestra estudiada.

Tamaño del grupo	Sabor	Otros atributos	Alcance
k = 3	Fresa, arándanos y coco	Light, Eco y 125 gr.	52,5 %
k = 4	Fresa, arándano, piña y coco	Light, Eco y 125 gr.	59,3 %
k = 5	Fresa, arándano, frambuesa, coco y <i>gin-tonic</i>	Light, Eco y 125 gr.	62,7 %
k = 6	Fresa, arándano, frambuesa, coco, piña y <i>gin-tonic</i>	Light, Eco y 125 gr.	66,1 %

Cuadro 5.2: Combinaciones resultantes para yogures

5.1.3. Perfiles de cremas de queso óptimos

Por último, y siguiendo los resultados de la sección 4.3.3, para ampliar la gama de cremas de queso disponibles para el consumidor, los resultados de la regresión logística indican que sería importante etiquetar el producto como ecológico, así como envasarlo en paquetes de 200 gramos. Queda descartado el etiquetado light, por no mejorar el ajuste del modelo. Así, este debería quedar sujeto a otros factores como los costes de producción o la imagen de la compañía, fuera del alcance de este estudio. Los sabores, de igual manera que se realizó para los dos productos anteriores, se escogen atendiendo a los resultados del análisis TURF, y en función del tamaño grupal que se quiera tomar. Están resumidos en el cuadro 5.3.

5.2. Críticas al trabajo

Una de las limitaciones más importantes, ya comentada a lo largo del texto, es el reducido tamaño de la muestra que se empleó para elaborar los modelos y análisis. Esto implica unos elevados errores estándar en el cálculo de los coeficientes de las variables significativas del modelo logístico (especialmente de aquellas menos populares), y un modelo que, en general, no se ajusta tan bien a los datos

Tamaño del grupo	Sabor	Otros atributos	Alcance
k = 3	Ahumado, yogur griego y vino de Oporto	Eco y 200 gr.	54,2%
k = 4	Ahumado, chocolate y naranja, yogur griego y <i>gin-tonic</i>	Eco y 200 gr.	61%
k = 5	Ahumado, chocolate y naranja, verduras a la brasa, yogur griego y vino de Oporto	Eco y 200 gr.	67,8%
k = 6	Jalapeño, ahumado, chocolate y naranja, verduras a la brasa y vino de Oporto	Eco y 200 gr.	72,9%

Cuadro 5.3: Combinaciones resultantes para cremas de queso

como sería deseable. Se hace más patente si cabe en la sección siguiente, en la que se trata de utilizar el modelo como un clasificador. Siguiendo con las limitaciones derivadas de la muestra, debido a que el método de muestreo carece de aleatoriedad, no se pueden extender las conclusiones a nivel poblacional, sino que la muestra tan sólo es representativa del círculo de allegados del autor.

Por otro lado, el autor pone en duda la calidad de los datos, al haber recibido reportes y comentarios acerca de la forma de completar la encuesta por parte de los encuestados. Así, algunos aseguran (especialmente entre los sectores de menor edad) que a partir de determinado punto de la encuesta, cuando se cansaron de prestar atención comenzaron a responder de forma más aleatoria. El hecho de que algunos individuos se hayan visto obligados a responder la encuesta varias veces, al haber sufrido problemas técnicos mientras respondían y verse obligados a volver a empezar, puede haber contribuido a este factor. Se especula sobre la aparente correlación inversa entre la edad y la pérdida del foco atencional. Aunque esta teoría contradice la literatura general sobre la atención sostenida (Vallesi et al. 2021), que indica que no aparecen grandes diferencias dependientes de la edad en tareas de atención sostenida y la fatiga que esta pudiera ocasionar, se estima que el contexto experimental pueda jugar un papel importante. Realizar la tarea en un laboratorio, bajo la presencia de un investigador y/o por una retribución económica puede elevar en gran medida la percepción de responsabilidad e importancia de la tarea a realizar, conllevando un esfuerzo consciente por parte del individuo para mantener la atención a lo largo de la misma y realizarla lo más correctamente posible. Sin embargo, el contexto experimental de una tarea no retribuida, con un entorno físico no controlado e interaccionando con el propio dispositivo (móvil o ordenador) podría reducir esa percepción de responsabilidad, siendo más elevada esta reducción en aquellos sujetos que acostumbran a pasar más horas delante de su dispositivo, conllevando una reducción de la atención prestada, de la importancia de la tarea percibida y consecuentemente una reducción en la calidad de los datos.

Apéndice A

Figuras

Helados

- Información demográfica
- Helados
- Yogures
- Creamas de queso

Información demográfica

Bienvenido/a y gracias por participar. Se te presenta una encuesta con cuatro apartados diferenciados: Información demográfica, Helados, Yogures y Creamas de queso. Son accesibles a través de la barra lateral. En el primero debes rellenar unos datos demográficos así como tu frecuencia de consumo de determinados productos lácteos.

En los apartados siguientes se mostrarán una selección de perfiles del mismo producto. Debes seleccionar la opción que te resulte más atractiva y pulsar continuar para guardar el resultado y acceder a la siguiente pregunta.

Puedes moverte libremente por la encuesta utilizando el panel lateral, pero no podrás cambiar una pregunta una vez que pulses continuar. Una vez que hayas contestado todas las preguntas de todos los apartados una nueva pantalla te dará las gracias por participar, en la que deberás pulsar un botón que dice guardar y salir.

Por favor indica tu edad

20

Por favor indica tu género

Masculino

Por favor indica la frecuencia con la que consumes helados en temporadas calurosas

Menos de una vez al mes

Por favor indica la frecuencia con la que consumes yogures

Menos de una vez al mes

Por favor indica la frecuencia con la que consumes cremas de queso

Menos de una vez al mes

Figura A.1: Interfaz del cuestionario, sección inicial

Helados

- Información demográfica
- Helados
- Yogures
- Creamas de queso

Helados

	Opción 1	Opción 2	Opción 3	Opción 4	Opción 5
Sabor	Vainilla	Plátano	Mango	Arándanos	Coco
Light	Si	No	Si	Si	Si
Presentación	Tarrina familiar	Ingrediente de un postre	Tarrina	Ingrediente de un postre	Tarrina
Eco	Si	Si	Si	No	No

La tabla muestra 5 helados distintos, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la forma de presentación y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva.

Señala tu opción preferida

Opción 1 Opción 2 Opción 3 Opción 4 Opción 5

Llevas 0 de 20 contestadas.

Figura A.2: Interfaz del cuestionario, sección helados

Helados

Información demográfica

Helados

Yogures

Cremas de queso

Yogures

	Opción 1	Opción 2	Opción 3	Opción 4	Opción 5
Sabor	Mango	Vainilla	Arándanos	Piña	Plátano
Light	No	Si	Si	Si	No
Cantidad	125 g	250 g	250 g	500 g	250 g
Eco	No	No	No	Si	Si

La tabla muestra 5 yogures distintos, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la cantidad que viene en el envase y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva

Señala tu opción preferida

Opción 1 Opción 2 Opción 3 Opción 4 Opción 5

Llevas 0 de 20 contestadas.

Continuar

Figura A.3: Interfaz del cuestionario, sección yogures

Helados

Información demográfica

Helados

Yogures

Cremas de queso

Cremas de queso

	Opción 1	Opción 2	Opción 3	Opción 4	Opción 5
Sabor	Chocolate y naranja	Vino de Oporto	Plátano	Jalapeño	Pimentón
Light	No	No	Si	Si	Si
Cantidad	500 g	500 g	500 g	500 g	200 g
Eco	Si	Si	No	Si	Si

La tabla muestra 5 cremas de queso distintas, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la cantidad que viene en el envase y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva

Señala tu opción preferida

Opción 1 Opción 2 Opción 3 Opción 4 Opción 5

Llevas 0 de 20 contestadas.

Continuar

Figura A.4: Interfaz del cuestionario, sección cremas de queso

Apéndice B

Código

B.1. Encuesta

```
library(shiny)
library(shinydashboard)
library(kableExtra)
library(rdrop2)
```

```
#####
```

```
##Generar elementos necesarios
```

```
intro <- "<p>Bienvenido/a y gracias por participar. Se te presenta una encuesta con cuatro apartados diferenciados: Información demográfica, Helados, Yogures y Cremas de queso. Son accesibles a través de la barra lateral</p>
```

```
<p>En el primero debes rellenar unos datos demográficos así como tu frecuencia de consumo de determinados productos lácteos.</p>
```

```
<p>En los apartados siguientes se mostrarán una selección de perfiles del mismo producto. Debes seleccionar la opción que te resulte más atractiva y pulsar continuar para guardar el resultado y acceder a la siguiente pregunta.</p>
```

```
<p>Puedes moverte libremente por la encuesta utilizando el panel lateral, pero no podrás cambiar una pregunta una vez que pulses continuar.
```

```
Una vez que hayas contestado todas las preguntas de todos los apartados una nueva pantalla te dará las gracias por participar, en la que deberás pulsar un botón que dice guardar y salir.</p>"
```

```
instr.breves.h <- "La tabla muestra 5 helados distintos, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la forma de presentación y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva"
```

```
instr.breves.y <- "La tabla muestra 5 yogures distintos, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la cantidad que viene en el envase y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva"
```

```
instr.breves.q <- "La tabla muestra 5 cremas de queso distintas, con cuatro atributos cada uno: Sabor, light (bajo en grasa), la cantidad que viene en el envase y eco (su producción es ecológica). Por favor selecciona la opción que te resulte más atractiva"
```

```
atrib.h <- list(Sabor = c("Fresa", "Vainilla", "Arándanos", "Mango", "Frambuesa",
```

```

"Plátano", "Maracuyá", "Coco", "Piña", "Gin-tonic",
"Manzana y canela", "Calabaza"),
  Light = c("Si", "No"),
  Presentacion = c("Cucurucho", "Tarrina", "Tarrina familiar", "Ingrediente
de un postre"),
  Eco = c("Si", "No")) #Todos los atributos de los helados

atrib.y <- list(Sabor = c("Fresa", "Vainilla", "Arándanos", "Mango", "Frambuesa",
"Plátano", "Maracuyá", "Coco", "Piña", "Gin-tonic"),
  Light = c("Si", "No"),
  Cantidad = c("125 g", "250 g", "500 g"),
  Eco = c("Si", "No"))

atrib.q <- list(Sabor = c("Jalapeño", "Ahumado", "Crema agria y cebolla", "Pimentón",
"Chocolate y naranja", "Verduras a la brasa", "Yogurt griego",
"Plátano", "Vino de Oporto"),
  Light = c("Si", "No"),
  Cantidad = c("200 g", "500 g"),
  Eco = c("Si", "No"))

perfiles.h <- expand.grid(atrib.h) #Todas las combinaciones de atributos
perfiles.y <- expand.grid(atrib.y)
perfiles.q<- expand.grid(atrib.q)
#Expand.grid no tiene aleatoriedad, no necesitamos fijar semilla

npreg.h <- 20 #número de preguntas a realizar
nalt.h <- 5 #nº de alternativas (perfiles por pregunta)
perfiles.seleccionables.h <- c("Opción 1", "Opción 2", "Opción 3",
"Opción 4", "Opción 5")
#esto con un paste0() quedaba mejor

npreg.y <- 20 #número de preguntas a realizar
nalt.y <- 5 #nº de alternativas (perfiles por pregunta)
perfiles.seleccionables.y <- c("Opción 1", "Opción 2", "Opción 3",
"Opción 4", "Opción 5")

npreg.q <- 20 #número de preguntas a realizar
nalt.q <- 5 #nº de alternativas (perfiles por pregunta)
perfiles.seleccionables.q <- c("Opción 1", "Opción 2", "Opción 3",
"Opción 4", "Opción 5")

preguntas.h <- list()
preguntas.y <- list()
preguntas.q <- list()

for (i in 1:npreg.h) {
  #set.seed(1299) #IMPORTANTE ALMOHADILLAR ESTO, LA TENGO PUESTA PARA
  #TESTEO, LUEGO LO QUEREMOS RANDOM
  ind.h <- sample(nrow(perfiles.h), nalt.h, replace = F)
  #índice para seleccionar los 5 perfiles
  pregx.h <- perfiles.h[ind.h,]
  rownames(pregx.h) <- perfiles.seleccionables.h

```

```

preguntas.h[[i]] <- t(pregx.h)
}
#el bucle devuelve 20 df (20 preguntas), con 5 perfiles cada una, en una lista
preguntas.h[[npreg.h+1]] <- preguntas.h[[npreg.h]]
#Duplicamos la última (la respuesta del duplicado ni se guarda) porque
#se bugea cuando el contador llega a 20 y no tiene otra tabla que renderizar

for (i in 1:npreg.y) {
  #set.seed(1299)
  #IMPORTANTE ALMOHADILLAR ESTO, LA TENGO PUESTA PARA TESTEO, LUEGO LO QUEREMOS RANDOM
  ind.y <- sample(nrow(perfiles.y), nalt.y, replace = F)
  #índice para seleccionar los 5 perfiles
  pregx.y <- perfiles.y[ind.y,]
  rownames(pregx.y) <- perfiles.seleccionables.y
  preguntas.y[[i]] <- t(pregx.y)
}
preguntas.y[[npreg.y+1]] <- preguntas.y[[npreg.y]]

for (i in 1:npreg.q) {
  #set.seed(1299)
  #IMPORTANTE ALMOHADILLAR ESTO, LA TENGO PUESTA PARA TESTEO, LUEGO LO QUEREMOS RANDOM
  ind.q <- sample(nrow(perfiles.q), nalt.q, replace = F)
  #índice para seleccionar los 5 perfiles
  pregx.q <- perfiles.q[ind.q,]
  rownames(pregx.q) <- perfiles.seleccionables.q
  preguntas.q[[i]] <- t(pregx.q)
}
preguntas.q[[npreg.q+1]] <- preguntas.q[[npreg.q]]

playerdata.h <- NULL
playerdata.y <- NULL
playerdata.q <- NULL

ques.h <- c()
ques.y <- c()
ques.q <- c()

for (i in 1:npreg.h) {
  ques.h <- c(ques.h, rep(i, nalt.h))
  playerdata.h <- rbind(playerdata.h, t(preguntas.h[[i]]))
}

for (i in 1:npreg.y) {
  ques.y <- c(ques.y, rep(i, nalt.y))
  playerdata.y <- rbind(playerdata.y, t(preguntas.y[[i]]))
}

for (i in 1:npreg.q) {
  ques.q <- c(ques.q, rep(i, nalt.q))
  playerdata.q <- rbind(playerdata.q, t(preguntas.q[[i]]))
}

```

```

playerdata.h <- cbind(ques.h, playerdata.h)
playerdata.y <- cbind(ques.y, playerdata.y)
playerdata.q <- cbind(ques.q, playerdata.q)
#Matrices de recogida de datos

#####
## Dropbox
droptoken <- readRDS("droptoken.rds")
#Token para la cuenta dropbox de jabkloinc@gmail.com
drop_acc(dtoken = droptoken)
#Si se roba este token se tiene acceso a este dropbox
outputDir <- "Resultados"
outputDir.h <- "Resultados.h"
outputDir.y <- "Resultados.y"
outputDir.q <- "Resultados.q"

#####
##APP
#####
ui <- dashboardPage(
  dashboardHeader(title = "Helados"),
  dashboardSidebar(
    sidebarMenu(id = "tabs",
      menuItem("Información demográfica", tabName = "demo", icon = icon("th")),
      menuItem("Helados", tabName = "cuerpoh", icon = icon("th")),
      menuItem("Yogures", tabName = "cuerpoy", icon = icon("th")),
      menuItem("Cremas de queso", tabName = "cuerpoq", icon = icon("th"))
    )),
  dashboardBody(
    tags$div(
      #En esta linea lo natural es tabItems() pero da bug con las tags de HTML.
      tabItem(tabName = "demo",h2("Información demográfica"),
        HTML(text = intro),
        numericInput("EDAD", "Por favor indica tu edad", value = 20, min = 5,
          max = 110, step = 1),
        selectInput("GENERO", "Por favor indica tu género", choices = c("Masculino",
          "Femenino", "No binario", "Prefiero no decirlo"), multiple=F),
        selectInput("FRECUENCIA.H", "Por favor indica la frecuencia con la que
          consumes helados en temporadas calurosas", choices = c("Menos de una vez
          al mes", "Menos de una vez a la semana", "Varias veces por semana"),
          multiple=F),
        selectInput("FRECUENCIA.Y", "Por favor indica la frecuencia con la que
          consumes yogures", choices = c("Menos de una vez al mes", "Menos de una
          vez a la semana", "Varias veces por semana"), multiple=F),
        selectInput("FRECUENCIA.Q", "Por favor indica la frecuencia con la que
          consumes cremas de queso", choices = c("Menos de una vez al mes", "Menos
          de una vez a la semana", "Varias veces por semana"), multiple=F)
      #,actionButton(inputId = 'switchtab.h', 'Ir a helados'),
      #actionButton(inputId = 'switchtab.q', 'Ir a yogures'),
      #actionButton(inputId = 'switchtab.y', 'Ir a cremas de queso')
      #Daban problema con las tags de html estos tres, los suprimí e indiqué que
      se navegue mediante la barra lateral
    )
  )
)

```

```

    ),
    tabItem(tabName = "cuerpoh",h2("Helados"), htmlOutput("table.h"),
      HTML(text = instr.breves.h),
      radioButtons("PH",label = "Señala tu opción preferida", choices =
        perfiles.seleccionables.h, inline = T),
      textOutput("progreso.h"),
      actionButton(inputId = 'continue.h', 'Continuar'),
      #verbatimTextOutput("debug.h") # print para debug
    ),
    tabItem(tabName = "cuerpoy",h2("Yogures"), htmlOutput("table.y"),
      HTML(text = instr.breves.y),
      radioButtons("PY",label = "Señala tu opción preferida", choices =
        perfiles.seleccionables.y, inline = T),
      textOutput("progreso.y"),
      actionButton(inputId = 'continue.y', 'Continuar'),
      #verbatimTextOutput("debug.y") # print para debug
    ),
    tabItem(tabName = "cuerpoq",h2("Cremas de queso"), htmlOutput("table.q"),
      HTML(text = instr.breves.q),
      radioButtons("PQ",label = "Señala tu opción preferida", choices =
        perfiles.seleccionables.q, inline = T),
      textOutput("progreso.q"),
      actionButton(inputId = 'continue.q', 'Continuar'),
      #verbatimTextOutput("debug.q") # print para debug
    ),
    class = "tab-content"
  ))
)

```

```
#####
```

```

server <- function(input, output, session) {

  # valores reactivos
  res.h <- reactiveVal("Perfiles:")
  res.y <- reactiveVal("Perfiles:")
  res.q <- reactiveVal("Perfiles:")
  contador <- reactiveVal(0) #Pequeño homenaje al ciclismo español
  indurain <- reactiveVal(0) #Indurain los yogures por el anuncio de danacol
  perico <- reactiveVal(0)

  #Los accesos a cada una de las pestañas desde la primera
  #observeEvent(input$switchtab.h, {
  # newtab <- switch(input$tabs, "demo" = "cuerpoh", "cuerpoh" = "demo")
  # updateTabItems(session, "tabs", newtab)
  #})
  #observeEvent(input$switchtab.y, {
  # newtab <- switch(input$tabs, "demo" = "cuerpoy", "cuerpoy" = "demo")
  # updateTabItems(session, "tabs", newtab)
  #})
  #observeEvent(input$switchtab.q, {
  # newtab <- switch(input$tabs, "demo" = "cuerpoq", "cuerpoq" = "demo")

```



```

#Comentado pues esto no debería mostrarse, solo es para el desarrollo de la app
# output$debug.h <- renderPrint({
#   list(res.h = res.h(),
#         contador = contador())
# })
#
# output$debug.y <- renderPrint({
#   list(res.y = res.y(),
#         indurain = indurain())
# })
#
# output$debug.q <- renderPrint({
#   list(res.q = res.q(),
#         perico = perico())
# })

observeEvent(input$continue.h, {
  #No se si este tramo, con tres observeEvent(uno por encuesta) y tres if()
  #dentro de cada uno sería optimizable. Probablemente

  if (contador() < length(preguntas.h)-1){
    # guardar resultados
    res.h(c(res.h(), input$PH))

    # contador y final
    contador(contador() + 1)
  }

  else if (contador() == length(preguntas.h)-1 & indurain() == length(preguntas.y)-1
  & perico() == length(preguntas.q)-1){
    showModal(modalFinal())
    edad <- input$EDAD
    genero <- input$GENERO
    frecuencia.h <- input$FRECUENCIA.H
    frecuencia.y <- input$FRECUENCIA.Y
    frecuencia.q <- input$FRECUENCIA.Q
    respuestas.h <- res.h()
    respuestas.y <- res.y()
    respuestas.q <- res.q()
    player_all <- list(playerdata.h, playerdata.y, playerdata.q, respuestas.h,
    respuestas.y, respuestas.q, edad, genero, frecuencia.h, frecuencia.y, frecuencia.q)
    # nombre del archivo
    filename <- paste0("resp", substring(session$token,1,8), ".rds")
    filepath <- file.path(tempdir(),filename)
    saveRDS(player_all, filepath)
    drop_upload(filepath, path = outputDir, dtoken = droptoken)
    # saveData(player_all)
  }

  else if (contador() == length(preguntas.h)-1){

```

```

#Solo enseña el modal, comentado para guardar resultados parciales en local.
#Online es un lío
edad <- input$EDAD
genero <- input$GENERO
frecuencia.h <- input$FRECUENCIA.H
respuestas.h <- res.h()
player_hel <- list(playerdata.h, respuestas.h, edad, genero, frecuencia.h)
filename <- paste0("resp.h", substring(session$token,1,8), ".rds")
filepath <- file.path(tempdir(),filename)
saveRDS(player_hel, filepath)
drop_upload(filepath, path = outputDir.h, dtoken = droptoken)
showModal(modalParcial())
}
})

observeEvent(input$continue.y, {

  if (indurain() < length(preguntas.y)-1){
    # guardar resultados
    res.y(c(res.y(), input$PY))

    # contador y final
    indurain(indurain() + 1)
  }

  else if (contador() == length(preguntas.h)-1 & indurain() == length(preguntas.y)-1 &
    perico() == length(preguntas.q)-1){
    showModal(modalFinal())
    edad <- input$EDAD
    genero <- input$GENERO
    frecuencia.h <- input$FRECUENCIA.H
    frecuencia.y <- input$FRECUENCIA.Y
    frecuencia.q <- input$FRECUENCIA.Q
    respuestas.h <- res.h()
    respuestas.y <- res.y()
    respuestas.q <- res.q()
    player_all <- list(playerdata.h, playerdata.y, playerdata.q, respuestas.h,
    respuestas.y, respuestas.q, edad, genero, frecuencia.h, frecuencia.y, frecuencia.q)
    filename <- paste0("resp", substring(session$token,1,8), ".rds")
    filepath <- file.path(tempdir(),filename)
    saveRDS(player_all, filepath)
    drop_upload(filepath, path = outputDir, dtoken = droptoken)
    # saveData(player_all)
    # saveData(player_all)
  }

  else if (indurain() == length(preguntas.y)-1){
    edad <- input$EDAD
    genero <- input$GENERO
    frecuencia.y <- input$FRECUENCIA.Y
    respuestas.y <- res.y()
    player_yor <- list(playerdata.y, respuestas.y, edad, genero, frecuencia.y)
  }
}

```

```

    filename <-paste0("resp.y", substring(session$token,1,8), ".rds")
    filepath <- file.path(tempdir(),filename)
    saveRDS(player_yor, filepath)
    drop_upload(filepath, path = outputDir.y, dtoken = droptoken)
    showModal(modalParcial())
  }
})

observeEvent(input$continue.q, {

  if (perico() < length(preguntas.q)-1){
    # guardar resultados
    res.q(c(res.q(), input$PQ))

    # contador y final
    perico(perico() + 1)
  }

  if (contador() == length(preguntas.h)-1 & indurain() == length(preguntas.y)-1 &
    perico() == length(preguntas.q)-1){
    showModal(modalFinal())
    edad <- input$EDAD
    genero <- input$GENERO
    frecuencia.h <- input$FRECUENCIA.H
    frecuencia.y <- input$FRECUENCIA.Y
    frecuencia.q <- input$FRECUENCIA.Q
    respuestas.h <- res.h()
    respuestas.y <- res.y()
    respuestas.q <- res.q()
    player_all <- list(playerdata.h, playerdata.y, playerdata.q, respuestas.h, respuestas.y,
      respuestas.q, edad, genero, frecuencia.h, frecuencia.y, frecuencia.q)
    filename <-paste0("resp", substring(session$token,1,8), ".rds")
    filepath <- file.path(tempdir(),filename)
    saveRDS(player_all, filepath)
    drop_upload(filepath, path = outputDir,dtoken = droptoken)
    # saveData(player_all)
    # saveData(player_all)
  }

  else if (perico() == length(preguntas.q)-1){
    edad <- input$EDAD
    genero <- input$GENERO
    frecuencia.q <- input$FRECUENCIA.Q
    respuestas.q <- res.q()
    player_que <- list (playerdata.q, respuestas.q, edad, genero, frecuencia.q)
    filename <-paste0("resp.q", substring(session$token,1,8), ".rds")
    filepath <- file.path(tempdir(),filename)
    saveRDS(player_que, filepath)
    drop_upload(filepath, path = outputDir.q, dtoken = droptoken)
    # saveRDS(player_all,"shiny2res.rds")
    showModal(modalParcial())
  }
}

```

```

})

#modal intermedio
modalParcial <- function(){
  modalDialog("Has acabado esta parte, por favor, navega por la barra lateral para
  finalizar el resto de apartados", easyClose = T, footer = NULL)
}

# modal ending

modalFinal <- function(){
  modalDialog("Has acabado, Gracias por colaborar!",
              footer = tagList(actionButton("endses","Guardar y salir"))
  )
}

observeEvent(input$endses,{
  Sys.sleep(1)
  stopApp()
})
}

shinyApp(ui = ui, server = server)

```

B.2. Preprocesado de los datos

```

#PREPROCESADO
#Data analysis
library(fastDummies)
library(dplyr)
setwd("C://Users/diego/Dropbox/Resultados")

#####
#Preprocesado de las respuestas a las 3 encuestas

#file path to storage
path <- "C://Users/diego/Dropbox/Resultados"
#Vector of al .rds files in folder
dir.vector <- dir(path, pattern = ".rds")

#Elementos que necesita el bucle
all_players <- list() #Storage superlist
opciones <- c("Opción 1", "Opción 2", "Opción 3", "Opción 4", "Opción 5") #For the IFs
long_format.h <- c()
long_format.y <- c()
long_format.q <- c()

#Loop para preprocesar los datos de cada encuestado
for (i in 1:length(dir.vector)){

```

```

print(i)
player_full <- readRDS(dir.vector[i])
#Separar elementos de la lista
playerdata.h <- player_full[[1]]
playerdata.y <- player_full[[2]]
playerdata.q <- player_full[[3]]
res.h <- player_full [[4]]
res.y <- player_full [[5]]
res.q <- player_full [[6]]
age <- player_full[[7]]
gen <- player_full [[8]]
freq.h <- player_full[[9]]
freq.y <- player_full[[10]]
freq.q <- player_full[[11]]

#Helados: codificacion dummie de la Y de los helados
res.h <- res.h[-1]
res.h <- c(opciones, res.h)
#Not the finnest way to get arround some pple never selecting one option
res.h.dum <- dummy_cols(res.h) #So dummy_cols() only created 4 columns instead of 5
res.h.dum <- res.h.dum[-(1:5),] #but it works and i'm fine with it
colnames(res.h.dum) <- NULL; rownames(res.h.dum) <- NULL
res.h.dum <- res.h.dum[,-1]
res.h.dum <- as.matrix(res.h.dum)
res.h.cod <- c()
for (j in 1:nrow(res.h.dum)){
  res.h.cod <- c(res.h.cod, res.h.dum[j,])
}
ID <- rep(i, nrow(playerdata.h))
nalt <- rep(1:5, 20)
edad <- rep(age, nrow(playerdata.h))
frec.h <- rep(freq.h, nrow(playerdata.h))
genero <- rep(gen, nrow(playerdata.h))
playerdata.h <- cbind(ID, nalt, playerdata.h, edad, genero, frec.h, res.h.cod)

#Yogures: codificacion dummie de los yogures
res.y <- res.y[-1]
res.y <- c(opciones, res.y)
res.y.dum <- dummy_cols(res.y)
res.y.dum <- res.y.dum[-(1:5),]
colnames(res.y.dum) <- NULL
res.y.dum <- res.y.dum[,2:6]
res.y.dum <- as.matrix(res.y.dum)
res.y.cod <- c()
for (k in 1:nrow(res.y.dum)){
  res.y.cod <- c(res.y.cod, res.y.dum[k,])
}
ID <- rep(i, nrow(playerdata.y))
nalt <- rep(1:5, 20)
edad <- rep(age, nrow(playerdata.y))
frec.y <- rep(freq.y, nrow(playerdata.y))
genero <- rep(gen, nrow(playerdata.y))

```

```

playerdata.y <- cbind(ID, nalt, playerdata.y, edad, genero, frec.y , res.y.cod)

#Quesos: codificacion dummie de los quesos
res.q <- res.q[-1]
res.q <- c(opciones, res.q)
res.q.dum <- dummy_cols(res.q)
res.q.dum <- res.q.dum[-(1:5),]
colnames(res.q.dum) <- NULL
res.q.dum <- res.q.dum[,2:6]
res.q.dum <- as.matrix(res.q.dum)
res.q.cod <- c()
for (l in 1:nrow(res.q.dum)){
  res.q.cod <- c(res.q.cod, res.q.dum[l,])
}
ID <- rep(i, nrow(playerdata.q))
nalt <- rep(1:5, 20)
edad <- rep(age, nrow(playerdata.q))
frec.q <- rep(freq.q, nrow(playerdata.q))
genero <- rep(gen, nrow(playerdata.q))
playerdata.q <- cbind(ID, nalt, playerdata.q, edad, genero, frec.q, res.q.cod)

#Reunite everything
player_full <- list(playerdata.h, playerdata.y, playerdata.q, edad, gen, frec.h, frec.y,
frec.q)

#Store in superlist
all_players[[i]] <- player_full

#Generate long format data frames
long_format.h <- rbind(long_format.h, playerdata.h)
long_format.y <- rbind(long_format.y, playerdata.y)
long_format.q <- rbind(long_format.q, playerdata.q)
}

long_format.h <- as.data.frame(long_format.h)
long_format.h[,1] <- as.numeric(long_format.h[,1])
long_format.h[,2] <- as.numeric(long_format.h[,2])
long_format.h[,3] <- as.numeric(long_format.h[,3])
long_format.h[,4] <- as.factor(long_format.h[,4])
long_format.h[,5] <- as.factor(long_format.h[,5])
long_format.h[,6] <- as.factor(long_format.h[,6])
long_format.h[,7] <- as.factor(long_format.h[,7])
long_format.h[,8] <- as.numeric(long_format.h[,8])
long_format.h[,9] <- as.factor(long_format.h[,9])
long_format.h[,10] <- as.factor(long_format.h[,10])
long_format.h[,11] <- as.factor(long_format.h[,11])
#long_format.h[,11] <- as.logical(as.integer(long_format.h[,11]))

long_format.h$$Sabor <- relevel(long_format.h$$Sabor, ref = "Fresa")
long_format.h$Light <- relevel(long_format.h$Light, ref = "No")
long_format.h$Presentacion <- relevel(long_format.h$Presentacion, ref = "Cucurucho")

```

```

long_format.h$Eco <- relevel(long_format.h$Eco, ref = "No")
long_format.h$genero <- relevel(long_format.h$genero, ref = "Masculino")
long_format.h$frec.h <- relevel(long_format.h$frec.h, ref="Varias veces por semana")

```

```

long_format.y <- as.data.frame(long_format.y)
long_format.y[,1] <- as.numeric(long_format.y[,1])
long_format.y[,3] <- as.numeric(long_format.y[,2])
long_format.y[,2] <- as.numeric(long_format.y[,3])
long_format.y[,4] <- as.factor(long_format.y[,4])
long_format.y[,5] <- as.factor(long_format.y[,5])
long_format.y[,6] <- as.factor(long_format.y[,6])
long_format.y[,7] <- as.factor(long_format.y[,7])
long_format.y[,8] <- as.numeric(long_format.y[,8])
long_format.y[,9] <- as.factor(long_format.y[,9])
long_format.y[,10] <- as.factor(long_format.y[,10])
long_format.y[,11] <- as.factor(long_format.y[,11])
#long_format.y[,11] <- as.logical(as.integer(long_format.y[,11]))

```

```

long_format.y$Sabor <- relevel(long_format.y$Sabor, ref= "Fresa")
long_format.y$Light <- relevel(long_format.y$Light, ref = "No")
long_format.y$Cantidad <- relevel(long_format.y$Cantidad, ref = "125 g")
long_format.y$Eco <- relevel(long_format.y$Eco, ref = "No")
long_format.y$genero <- relevel(long_format.y$genero, ref = "Masculino")
long_format.y$frec.y <- relevel(long_format.y$frec.y, ref="Varias veces por semana")

```

```

long_format.q <- as.data.frame(long_format.q)
long_format.q[,1] <- as.numeric(long_format.q[,1])
long_format.q[,2] <- as.numeric(long_format.q[,2])
long_format.q[,3] <- as.numeric(long_format.q[,3])
long_format.q[,4] <- as.factor(long_format.q[,4])
long_format.q[,5] <- as.factor(long_format.q[,5])
long_format.q[,6] <- as.factor(long_format.q[,6])
long_format.q[,7] <- as.factor(long_format.q[,7])
long_format.q[,8] <- as.numeric(long_format.q[,8])
long_format.q[,9] <- as.factor(long_format.q[,9])
long_format.q[,10] <- as.factor(long_format.q[,10])
long_format.q[,11] <- as.factor(long_format.q[,11])
#long_format.q[,11] <- as.logical(as.integer(long_format.q[,11]))

```

```

long_format.q$Sabor <- relevel(long_format.q$Sabor, ref = "Pimentón")
long_format.q$Light <- relevel(long_format.q$Light, ref = "No")
long_format.q$Cantidad <- relevel(long_format.q$Cantidad, ref = "200 g")
long_format.q$Eco <- relevel(long_format.q$Eco, ref = "No")
long_format.q$genero <- relevel(long_format.q$genero, ref = "Masculino")
long_format.q$frec.q <- relevel(long_format.q$frec.q, ref="Varias veces por semana")

```

```

setwd("c:/datos/Encuesta_lacteos/choice-based_conjoint")
save(all_players, long_format.h, long_format.y, long_format.q,
file = "lacteos_data_long.RData")

```

###TURF PREPROCESSING

```

atrib.h <- list(Sabor = c("Fresa", "Vainilla", "Arándanos", "Mango", "Frambuesa", "Plátano",
  "Maracuyá", "Coco", "Piña", "Gin-tonic", "Manzana y canela",
  "Calabaza"),
  Light = c("Si", "No"),
  Presentacion = c("Cucurucho", "Tarrina", "Tarrina familiar", "Ingrediente de
    un postre"),
  Eco = c("Si", "No")) #Todos los atributos de los helados

atrib.y <- list(Sabor = c("Fresa", "Vainilla", "Arándanos", "Mango", "Frambuesa", "Plátano",
  "Maracuyá", "Coco", "Piña", "Gin-tonic"),
  Light = c("Si", "No"),
  Cantidad = c("125 g", "250 g", "500 g"),
  Eco = c("Si", "No"))

atrib.q <- list(Sabor = c("Jalapeño", "Ahumado", "Crema agria y cebolla", "Pimentón",
  "Chocolate y naranja", "Verduras a la brasa", "Yogurt griego",
  "Plátano", "Vino de Oporto"),
  Light = c("Si", "No"),
  Cantidad = c("200 g", "500 g"),
  Eco = c("Si", "No"))

perfiles.h <- expand.grid(atrib.h) #Todas las combinaciones de atributos
perfiles.y <- expand.grid(atrib.y)
perfiles.q <- expand.grid(atrib.q)

perfiles.h$perfil <- rownames(perfiles.h)
perfiles.y$perfil <- rownames(perfiles.y)
perfiles.q$perfil <- rownames(perfiles.q)

preturfddata.h <- merge(long_format.h, perfiles.h)
preturfddata.h$res.h.cod <- ifelse(preturfddata.h$res.h.cod==0,0,1)
preturfddata.y <- merge(long_format.y, perfiles.y)
preturfddata.y$res.y.cod <- ifelse(preturfddata.y$res.y.cod==0,0,1)
preturfddata.q <- merge(long_format.q, perfiles.q)
preturfddata.q$res.q.cod <- ifelse(preturfddata.q$res.q.cod==0,0,1)

ID <- seq(1,59,1)
weight <- rep(1,59)

perf.h <- seq(1,192,1)
turfddata.h <- matrix(nrow = 59, ncol = 194)
turfddata.h[,1] <- ID; turfddata.h[,2] <- weight
colnames(turfddata.h) <- c("ID", "weight", paste0("p",perf.h))

perf.y <- seq(1,120,1)
turfddata.y <- matrix(nrow = 59, ncol = 122)
turfddata.y[,1] <- ID; turfddata.y[,2] <- weight
colnames(turfddata.y) <- c("ID", "weight", paste0("p",perf.y))

```

```

perf.q <- seq(1,72,1)
turfddata.q <- matrix(nrow = 59, ncol = 74)
turfddata.q[,1] <- ID; turfddata.q[,2] <- weight
colnames(turfddata.q) <- c("ID", "weight", paste0("p",perf.q))

for (i in 1:59) {
  print(i)
  a <- preturfddata.h %>% filter(ID==i)
  for (j in 1:192){
    b <- a %>% filter(perfil == j)
    print(j)
    if (nrow(b)==1)
      turfddata.h[i,j+2] <- ifelse(b$res.h.cod==1,1,0)
    else if(nrow(b)==0)
      turfddata.h[i,j+2] <- 0
    else if (nrow(b)>1)
      turfddata.h[i,j+2] <- ifelse(sum(b$res.h.cod)>0,1,0)
  }
}
turfddata.h <- as.data.frame(turfddata.h)

for (i in 1:59) {
  print(i)
  a <- preturfddata.y %>% filter(ID==i)
  for (j in 1:120){
    b <- a %>% filter(perfil == j)
    print(j)
    if (nrow(b)==1)
      turfddata.y[i,j+2] <- ifelse(b$res.y.cod==1,1,0)
    else if(nrow(b)==0)
      turfddata.y[i,j+2] <- 0
    else if (nrow(b)>1)
      turfddata.y[i,j+2] <- ifelse(sum(b$res.y.cod)>0,1,0)
  }
}
turfddata.y <- as.data.frame(turfddata.y)

for (i in 1:59) {
  print(i)
  a <- preturfddata.q %>% filter(ID==i)
  for (j in 1:72){
    b <- a %>% filter(perfil == j)
    print(j)
    if (nrow(b)==1)
      turfddata.q[i,j+2] <- ifelse(b$res.q.cod==1,1,0)
    else if(nrow(b)==0)
      turfddata.q[i,j+2] <- 0
    else if (nrow(b)>1)
      turfddata.q[i,j+2] <- ifelse(sum(b$res.q.cod)>0,1,0)
  }
}

```

```

    }
  }
  turfdata.q <- as.data.frame(turfdata.q)

  setwd("c:/datos/Encuesta_lacteos/choice-based_conjoint")
  save(turfdata.h, turfdata.y, turfdata.q, file = "lacteos_turfdata.RData")

#REDUX TURF DATA
#'cause problem was too large

Sabor.h = c("Fresa", "Vainilla", "Arándanos", "Mango", "Frambuesa", "Plátano", "Maracuyá",
            "Coco", "Piña", "Gin-tonic", "Manzana y canela", "Calabaza")

Sabor.y = c("Fresa", "Vainilla", "Arándanos", "Mango", "Frambuesa", "Plátano", "Maracuyá",
            "Coco", "Piña", "Gin-tonic")

Sabor.q = c("Jalapeño", "Ahumado", "Crema agria y cebolla", "Pimentón", "Chocolate y naranja",
            "Verduras a la brasa", "Yogurt griego", "Plátano", "Vino de Oporto")

ID <- seq(1,59,1)
weight <- rep(1,59)

perf.h <- 1:length(Sabor.h)
turfdata.redux.h <- matrix(nrow = 59, ncol = length(perf.h)+2)
turfdata.redux.h[,1] <- ID; turfdata.redux.h[,2] <- weight
colnames(turfdata.redux.h) <- c("ID", "weight", Sabor.h)

perf.y <- 1:length(Sabor.y)
turfdata.redux.y <- matrix(nrow = 59, ncol = length(perf.y)+2)
turfdata.redux.y[,1] <- ID; turfdata.redux.y[,2] <- weight
colnames(turfdata.redux.y) <- c("ID", "weight", Sabor.y)

perf.q <- 1:length(Sabor.q)
turfdata.redux.q <- matrix(nrow = 59, ncol = length(perf.q)+2)
turfdata.redux.q[,1] <- ID; turfdata.redux.q[,2] <- weight
colnames(turfdata.redux.q) <- c("ID", "weight", Sabor.q)

for (i in ID){
  print(i)
  a <- preturfdata.h %>% filter(ID == i)
  for (j in perf.h){
    print(j)
    b <- a %>% filter(Sabor == Sabor.h[j])
    if (nrow(b)==1)
      turfdata.redux.h[i,j+2] <- ifelse(b$res.h.cod==1,1,0)
    else if(nrow(b)==0)
      turfdata.redux.h[i,j+2] <- 0
    else if (nrow(b)>1)
      turfdata.redux.h[i,j+2] <- ifelse(mean(b$res.h.cod)>=0.5,1,0)
  }
}
}

```

```

turfddata.redux.h <- as.data.frame(turfddata.redux.h)

for (i in ID){
  print(i)
  a <- preturfddata.y %>% filter(ID == i)
  for (j in perf.y){
    print(j)
    b <- a %>% filter(Sabor == Sabor.y[j])
    if (nrow(b)==1)
      turfddata.redux.y[i,j+2] <- ifelse(b$res.y.cod==1,1,0)
    else if(nrow(b)==0)
      turfddata.redux.y[i,j+2] <- 0
    else if (nrow(b)>1)
      turfddata.redux.y[i,j+2] <- ifelse(mean(b$res.y.cod)>=0.5,1,0)
  }
}
turfddata.redux.y <- as.data.frame(turfddata.redux.y)

for (i in ID){
  print(i)
  a <- preturfddata.q %>% filter(ID == i)
  for (j in perf.q){
    print(j)
    b <- a %>% filter(Sabor == Sabor.q[j])
    if (nrow(b)==1)
      turfddata.redux.q[i,j+2] <- ifelse(b$res.q.cod==1,1,0)
    else if(nrow(b)==0)
      turfddata.redux.q[i,j+2] <- 0
    else if (nrow(b)>1)
      turfddata.redux.q[i,j+2] <- ifelse(mean(b$res.q.cod)>=0.5,1,0)
  }
}
turfddata.redux.q <- as.data.frame(turfddata.redux.q)

setwd("c:/datos/Encuesta_lacteos/choice-based_conjoint")
save(turfddata.redux.h, turfddata.redux.y, turfddata.redux.q,
file = "lacteos_turfddata_redux.RData")

```

B.3. Análisis de los datos

Se avisa al lector de que, si bien este código corre entero sin problemas en la versión 4.0.1 de R siempre que se tengan los paquetes requeridos instalados y los datos que requiere, todo el proceso *stepwise* realizado por la librería *caret* puede llevar un tiempo considerable, pues son muchos los modelos diferentes calculados.

```

setwd("c:/datos/Encuesta_lacteos/choice-based_conjoint")

load("lacteos_data_long.RData") #DATA
load("lacteos_turfddata.RData") #TURF DATA
load("lacteos_turfddata_redux.RData") #TURF DATA PERO SOLO SABORES
load("modeloslogit59resp.RData") #LOGIT MODELS 59 answers

```

```
#####
##Libraries

library(dplyr)
library(ggplot2)
library(caret)
library(pROC)
library(xtable)
library(turfR)

#####
##Descriptive
edad <- c(); gen <- c(); frec.h <- c(); frec.y <- c(); frec.q <- c()
for (i in 1:59){
  playeri <- all_players[[i]]
  edad <- c(edad, as.numeric(playeri[[4]][1]))
  gen <- c(gen, playeri[[5]][1])
  frec.h <- c(frec.h, playeri[[6]][1])
  frec.y <- c(frec.h, playeri[[7]][1])
  frec.q <- c(frec.q, playeri[[8]][1])
}

summary(edad)

gen.cod <- c()
for (i in 1:length(gen)){
  if (gen[i] == "Femenino") {a=0}
  else if (gen[i] == "Masculino") {a=1}
  else if (gen[i] == "No binario") {a=2}
  else {a=3}
  gen.cod <- c(a, gen.cod)
}
summary(gen)

gen <- ifelse(gen=="Femenino",0,1)
gender <- factor(x=gen, levels = c(0,1,2,3),
                labels = c("Femenino", "Masculino", "No binario", "NS/NC"))
summary(gender)
barplot(table(gender))

frec.h.gg <- factor(frec.h, levels=c("Varias veces por semana", "Menos de una vez a la
semana", "Menos de una vez al mes"))
ggplot() + aes(frec.h.gg)+ geom_histogram(stat="count", colour="black", fill="white") +
  labs(x = "Frecuencia de consumo de helados", y = "nº de individuos") +
  scale_x_discrete(breaks = c(1,2,3), labels = c("< 1/mes", "< 1 vez/semana",
"> 1 vez/semana"))+
  theme(axis.title.x = element_text(size = 17),
        axis.title.y = element_text(size = 17))

frec.y.gg <- factor(frec.y, levels=c("Varias veces por semana", "Menos de una vez a la
semana", "Menos de una vez al mes"))
```

```

ggplot() + aes(frec.y.gg)+ geom_histogram(stat="count", colour="black", fill="white")+
  labs(x = "Frecuencia de consumo de yogures", y = "nº de individuos") +
  scale_x_discrete(breaks = c(1,2,3), labels = c("< 1/mes", "< 1 vez/semana",
    "> 1 vez/semana"))+
  theme(axis.title.x = element_text(size = 17),
    axis.title.y = element_text(size = 17))

frec.q.gg <- factor(frec.q, levels = c("Varias veces por semana", "Menos de una vez a la
semana", "Menos de una vez al mes"))
ggplot() + aes(frec.q.gg) +
  geom_histogram(stat="count", colour="black", fill="white") +
  labs(x = "Frecuencia de consumo de cremas de queso", y = "nº de individuos") +
  scale_x_discrete(breaks = c(1,2,3), labels = c("< 1/mes", "< 1 vez/semana",
    "> 1 vez/semana"))+
  theme(axis.title.x = element_text(size = 17),
    axis.title.y = element_text(size = 17))

##XTABS

long_format.h$res.h.cod <- as.numeric(long_format.h$res.h.cod)
long_format.h$res.h.cod <- ifelse(long_format.h$res.h.cod==1,0,1)
xtable(t(xtabs(res.h.cod ~ Sabor, data = long_format.h)))
xtable(t(xtabs(res.h.cod ~ Light, data = long_format.h))) #pa na casi
xtable(t(xtabs(res.h.cod ~ Presentacion, data = long_format.h))) #pa na casi
xtable(t(xtabs(res.h.cod ~ Eco, data = long_format.h))) #unos pocos más sies que noes

long_format.y$res.y.cod <- as.numeric(long_format.y$res.y.cod)
long_format.y$res.y.cod <- ifelse(long_format.y$res.y.cod==1,0,1)
#R guarda el factor como 1 y 2, al convertir a numerico se lía y luego las xtabs dan mal
xtable(t(xtabs(res.y.cod ~ Sabor, data = long_format.y)))
xtable(t(xtabs(res.y.cod ~ Light, data = long_format.y)))
xtable(t(xtabs(res.y.cod ~ Cantidad, data = long_format.y)))
xtable(t(xtabs(res.y.cod ~ Eco, data = long_format.y)))

long_format.q$res.q.cod <- as.numeric(long_format.q$res.q.cod)
long_format.q$res.q.cod <- ifelse(long_format.q$res.q.cod==1,0,1)
xtable(t(xtabs(res.q.cod ~ Sabor, data = long_format.q)))
xtable(t(xtabs(res.q.cod ~ Light, data = long_format.q)))
xtable(t(xtabs(res.q.cod ~ Cantidad, data = long_format.q)))
xtable(t(xtabs(res.q.cod ~ Eco, data = long_format.q)))

#####
###HELADOS
#####
###LOGIT

summary(long_format.h)
#Ponemos de referencia el helado de fresa, en cucurucho, ni light ni eco.

```

```

#Es como el mas corriente

##GLM Rbase
glm.h <- glm(res.h.cod ~ Sabor + Light + Presentacion + Eco, family = binomial(link = logit),
            data = long_format.h)

glm.i.h <- glm(res.h.cod ~ Sabor + Light + Presentacion + Eco + Eco*Presentacion,
            family = binomial(link=logit), data=long_format.h)
            #se comprobaron también las demás interacciones posibles

glm.null.h <- glm(res.h.cod ~ 1, family = binomial(link = logit),
            data = long_format.h)

# glm.h_0 <- glm(res.h.cod ~0+ Sabor + Light + Presentacion + Eco,
# family = binomial(link = logit), data = long_format.h)

summary(glm.h)
glm.step.h <-step(glm.h)
summary(glm.step.h)
exp(coef(glm.step.h))

anova(glm.step.h, glm.null.h, test="Chi") #Vale cunde más que el neutro dice

##GLM CARET
set.seed(8822)

ctrl <- trainControl(method = "cv",
                    number = 10,
                    selectionFunction = "oneSE")

caret.glm.h <- train(res.h.cod ~ Sabor + Light + Presentacion + Eco + edad + genero + frec.h,
                    data = train.h, method = "glmStepAIC", family = "binomial",
                    trControl = ctrl)

# caret.glm.h.null <- train(res.h.cod ~ 0, data = train.h, method = "glm",
# family = "binomial", trainControl = ctrl)

summary(caret.glm.h)

#####
###YOGURES
#####
##LOGIT

summary(long_format.y)

glm.y <- glm(res.y.cod ~ Sabor + Light + Cantidad + Eco,
            family = binomial(link = logit), data = long_format.y)

glm.i.y <- glm(res.y.cod ~ Sabor + Light + Cantidad + Eco + Eco*Cantidad,

```

```

        family = binomial(link = logit), data = long_format.y)

glm.null.y <- glm(res.y.cod ~ 1, family = binomial(link = logit), data = long_format.y)
summary(glm.y)
glm.step.y <- step(glm.y) #Ojo que no saca ninguna, aquí que sea light si que importa
summary(glm.step.y)
exp(coef(glm.step.y))
exp(confint(glm.step.y))

anova(glm.step.y, glm.null.y, test = "Chi")

##GLM CARET

ctrl <- trainControl(method = "cv",
                     number = 10,
                     selectionFunction = "oneSE")

caret.glm.y <- train(res.y.cod ~ Sabor + Light + Cantidad + Eco + edad + genero + frec.y,
                   data = train.y, method = "glmStepAIC", family = "binomial",
                   trControl = ctrl)

summary(caret.glm.y)

###QUESO
#####
##LOGIT

summary(long_format.q)

glm.q <- glm(res.q.cod ~ Sabor + Light + Cantidad + Eco + edad + genero + frec.q,
            family = binomial(link = logit), data = long_format.q)

glm.i.q <- glm(res.q.cod ~ Sabor + Light + Cantidad + Eco + Eco*Cantidad,
              family = binomial(link = logit), data = long_format.q)

glm.null.q <- glm(res.q.cod ~ 1, family = binomial(link = logit),
                 data = long_format.q)

glm.step.q <- step(glm.q)
summary(glm.step.q)
exp(coef(glm.step.q))
exp(confint(glm.step.q))

anova(glm.step.q, glm.null.q, test = "Chi")

#glm.q_0 <- glm(res.q.cod ~ 0 + Sabor + Light + Cantidad + Eco,
# family = binomial(link = logit), data = long_format.q)

```

```
#summary(glm.q_0)

#####
###TURF
###HELADOS

# turf(turfddata.h, 192, 3)
# turf(turfddata.h, 120, 3)
# turf(turfddata.h, 72, 3) Demasiado grande como para procesarlo.
#Vamos a probar con solo por sabores

##TODOS SIMPLIFICADOS (solo sabores)
turf(turfddata.redux.h, 12, 3:6)
turf(turfddata.redux.y, 10, 3:6)
turf(turfddata.redux.q, 9, 3:6)
```

B.4. Gráficos

```
library(eulerr)
data <- c(LVDG = 200,
         EP = 120,
         EFDV = 166,
         ABC = 87,
         "LVDG&EP" = 18,
         "LVDG&EFDV" = 140,
         "EP&EFDV" = 39)
venn <- euler(data)
plot(venn)
```

Bibliografía

- [1] Goldstein BE (2009). *Sensation and Perception* (8ed). Cenage Learning, Wadsworth.
- [2] Meilgaard MC, Civille GV y Carr BT (2007). *Sensory Evaluation Techniques* (6ed). CRC Press, Boca Ratón.
- [3] Issanchou S y Nicklaus S (2006). 6 - Measuring consumersperceptions of sweet taste. En: Spillane WJ, (1ed) *Optimising Sweet Taste in Foods*. Woodhead Publishing, pp 97-131.
- [4] Tomlison V (2019). On the Sensorial of Imagination. En: Coessens, K (1ed) *Sensorial Aesthetics in Music Practices*. Orpheus. pp 169-181
- [5] Zwerina K (1997). *Discrete Choice Experiments in Marketing: Use of Priors in Efficient Choice Designs and Their Application to Individual Preference Measurement*. Physica-Verlag HD, Heidelberg.
- [6] Chapman C, McDonell FE (2015). *R for Marketing Research and Analytics* (2ed).Springer Nature Switzerland, Cham.
- [7] Miaoulis G, Free V, Parsons H (1990) Turf: A New Planning Approach for Product Line Extensions. *Marketing Research*2:28-40.
- [8] Lipovetsky S (2008) SURF - Structural unduplicated reach and frequency: latent class TURF and Shapley value analyses. *International Journal of Information Technology & Decision Making* 7:203-216.
- [9] Conklin M, Lipovetsky S (1999) A Winning Tool for CPG. *Marketing Research* 11:22-27.
- [10] Kreiger AM, Green PE (2000)TURF Revisited: Enhancements to Total Unduplicated Reach and Frequency Analysis. *Marketing Research*, 12(4).
- [11] Carlson C, Little JM (2019). Identifying core themes for an Integrated HSEM curriculum. *International Journal of Security, Preparedness, and Resilience Education*. 1:1-22. <https://jsire.org/identifying-core-themes-for-an-integrated-hsem-curriculum>, <https://jsire.org/identifying-core-themes-for-an-integrated-hsem-curriculum>
- [12] Adler T, Smith C, Dumont J (2010). Optimizing Product Portfolios Using Discrete Choice Modeling and TURF. En: Hess S, Daly A (2010). *Choice Modeling: The State of the Art and the State of practice*. Emerald Group Publishing Limited, Bingley, pp. 483-497.
- [13] Ankur J, Saket K, Satish C (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*. 7:396-403
- [14] R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

- [15] Wolke R, Schwetlick H (1988). Iterative Reweighted Least Squares: Algorithms, Convergence Analysis, and Numerical Comparisons. *SIAM Journal on Scientific and Statistical Computing*. 9:907-921
- [16] Cluster Alimentario de Galicia (2020). Tendencias Globales Sector Lacteo 2020.
- [17] Sawtooth Studio (2021). Lighthouse Studio Overview. Sawtooth Software. <https://sawtoothsoftware.com/lighthouse-studio>. Accedido 27 de junio de 2021
- [18] Gogan LM, Duran DC, Anca D (2015). Structural Capital - A Proposed Measurement Model. *Procedia Economics and Finance*. 23:1139-1146
- [19] Bornstein MC, Jager J, Putnick DL (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review*. 33:357-370.
- [20] Vallesi A, Tronelli V, Lomi F, Pezzetta R (2021). Age differences in sustained attention tasks: A meta-analysis. *Psychonomic Bulletin & Review*. 28:1755-1775.
- [21] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, Jonathan McPherson (2020). shiny: Web Application. Framework for R. R package version 1.5.0. <https://CRAN.R-project.org/package=shiny>
- [22] Winston Chang and Barbara Borges Ribeiro (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. <https://CRAN.R-project.org/package=shinydashboard>
- [23] Hao Zhu (2020). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.1. <https://CRAN.R-project.org/package=kableExtra>
- [24] Karthik Ram and Clayton Yochum (2020). rdrop2: Programmatic Interface to the 'Dropbox' API. R package version 0.8.2.1. <https://CRAN.R-project.org/package=rdrop2>
- [25] Jack Horne (2014). turfR: TURF Analysis for R. R package version 0.8-7. <https://CRAN.R-project.org/package=turfR>
- [26] caret2020) Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. <https://CRAN.R-project.org/package=caret>
- [27] David B. Dahl, David Scott, Charles Roosen, Arni Magnusson and Jonathan Swinton (2019). xtable: Export Tables to LaTeX or HTML. R package version 1.8-4. <https://CRAN.R-project.org/package=xtable>
- [28] Larsson J (2020). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.1.0. <https://cran.r-project.org/package=eulerr>.
- [29] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. <https://ggplot2.tidyverse.org>