
Análisis de datos funcionales dispersos para señales electromiográficas (EMG)

Analysis of sparse functional data for electromyographic signals

Alejandro Lopera Marín.^a
alejandrolopera@usantotomas.edu.co

Wilmer Pineda Ríos.^b
wilmerpineda@usantotomas.edu.co

Carolina Garzón Medina.^c
carolinagarzonm@usantotomas.edu.co

Resumen

En este documento se plantea una metodología estructurada para el análisis de datos electromiográficos (EMG) por medio del análisis de datos funcionales para datos dispersos, que busca modelar y analizar el comportamiento de consumo de los individuos de bebidas lácteas saludables en una investigación desarrollada por el departamento de Mercadeo de la Universidad Santo Tomás. Los resultados obtenidos en esta investigación muestran la posibilidad de interpretar las señales EMG con la ayuda de un cuestionario, y así poder identificar los patrones de señales que indiquen el nivel de agrado o desagrado de los estímulos presentados para llegar a conclusiones que permitan identificar la reacción que tiene un individuo frente a un determinado estímulo.

Palabras Clave:: Datos dispersos, Análisis de Componentes Principales Funcionales, Análisis de conglomerados funcionales.

Abstract

This document proposes a structured methodology for the analysis of electromyographic data (EMG) through the analysis of functional data for sparse data, which seeks to model and analyze the consumption behavior of healthy milk drinkers in a research developed by the marketing department of the Santo Tomás University. The results in this investigation show the possibility of interpreting the EMG signals with the help of a questionnaire, and thus be able to identify the patterns of the signals that indicate the level of liking and the displeasure of the stimuli to reach the conclusions that can identify the reaction that an individual has against a certain stimulus.

Key Words: Sparses data, Functional Principal Component Analysis, functional conglomerate analysis

^aEstudiante pregrado en Estadística, U. Santo Tomás, sede Bogotá

^bProfesor pregrado en Estadística, U. Santo Tomás, sede Bogotá

^cProfesora pregrado en Mercadeo, U. Santo Tomás, sede Bogotá

1. Introducción

Durante la última década se ha venido desarrollando importantes investigaciones en temas relacionados con la neurociencia cognitiva aplicada al marketing, con el propósito de comprender de qué forma los consumidores perciben, valoran y prefieren los productos existentes en el mercado como una forma alternativa a la investigación tradicional y a partir de esto, diseñar, perfeccionar o reformular productos, servicios y marcas que tengan éxito en el mercado que progresivamente es más saturado y globalizado.

Importantes marcas a nivel mundial como General Electric (GE) y empresas dedicadas a la investigación de mercados como A.C. Nielsen y Forbes Consulting Group han mostrado gran interés en desarrollar centros de investigación dedicados a la neurociencia del consumidor, diseñando herramientas y metodologías que permitan caracterizar el comportamiento de consumo, esto con el propósito de corregir las deficiencias existentes en las metodologías tradicionales para obtener datos derivados de las encuestas y métodos cualitativos.

El diseño de estas nuevas herramientas y metodologías, como es el caso puntual del uso de la electromiografía (EMG), que permite medir la reacción eléctrica facial a partir estímulos presentados, ya sean visuales o auditivos, hace que sea indispensable aplicar métodos estadísticos bien estructurados que permitan modelar la complejidad propia de este tipo de datos, para poder llegar a obtener conclusiones más confiables.

A partir de lo anterior, este trabajo se basa en la construcción de una metodología estadística basada en el análisis de componentes principales para datos funcionales, en el que paso a paso se va desarrollando una metodología de análisis que muestra cómo procesar y analizar las complejas ondas EMG, comenzando por comprender qué son los datos funcionales, cómo a partir de datos discretizados se pueden llegar a datos funcionales, y para este caso en particular, en donde las frecuencias de tiempo medidas en los individuos no fue en la misma longitud, se puedan analizar por métodos de análisis de datos funcionales para datos dispersos, para finalmente por el método de conglomerados funcionales caracterizar el comportamiento de las ondas EMG.

2. Datos Funcionales

En estadística, datos funcionales se le llama al análisis de todos los datos que proporciona información en forma de curva u onda en un rango de tiempo determinado $T = [a, b] \subset \mathbb{R}$, que varía de forma continua, en donde cada registro se considera como una función (Ramsay., 2005). Algunos ejemplos de datos funcionales son, la temperatura de un país desde enero a diciembre durante los últimos 25 años, donde cada año es una función aleatoria, otro ejemplo es el caso de las señales obtenidas por electromiografía (EMG) como es caso que se analiza en esta investigación, en donde el impulso eléctrico generado por la expresión facial del sujeto es transformada e interpretada como una onda, que para este caso, cada estímulo presentado al sujeto, genera una señal en forma de onda que es considera como una función.

Como lo define (Ferrety and Vieu., 2006) una variable aleatoria \mathbf{X} , es aquella variable que toma valores en un espacio dimensional infinito (o espacio funcional) y una observación x de \mathbf{X} es llamado dato funcional.

$$\begin{aligned} \mathbf{X} : T &\longrightarrow \mathbb{R} \\ t &\longmapsto \mathbb{R} \end{aligned}$$

Si $\{x_1(t), x_2(t), \dots, x_n(t)\}$ es una m.a (iid), es observada de la forma $x = \{x(t); t \in T\}$, que para este caso, la característica funcional proviene de las observaciones, para el caso cuando la variable es una

curva que está asociada con un conjunto unidimensional y que es representado como un conjunto de pares finitos (t_i, x_i) $t_i \in T$, con $i = 1, 2, \dots, N$, donde N está representado por la cantidad de puntos que se encuentran dentro de la variable funcional de interés.

El espacio $L^2 = L^2(T)$ es el conjunto de funciones medibles de valor real X definido en T , que satisface:

$$L^2(T) = \left\{ f : T \rightarrow \mathbb{R} \mid \int_a^b f(t)^2 dt < \infty \right\}$$

El espacio L^2 es un espacio separable de Hilbert con el producto interno (Horváth and Kokoszka., 2012).

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt$$

Con la anterior definición, se garantiza que el conjunto discreto de datos se puede aproximar funciones que puedan suavizar el comportamiento de los datos, que para este caso, al trabajar con datos de electromiografía, la aproximación más útil para este caso es por medio de las series de Fourier.

2.1. Suavizado

Con el propósito de poder transformar los datos electromiográficos obtenidos en el experimento en datos funcionales, es necesario realizar una aproximación a una función $f(t)$ por medio de combinación lineal de funciones. Teniendo en cuenta la relación entre frecuencia y amplitud, la mejor forma de representar los datos es por medio de series de Fourier, de esta forma se logra a partir de un conjunto de datos discretos, obtener una buena aproximación a datos funcionales $f(t)$ que se expresa de la siguiente manera:

$$f(t) \approx \frac{a_0}{2} + \sum_{i=1}^N \left\{ a_i \cos\left(\frac{2\pi it}{N}\right) + b_i \sin\left(\frac{2\pi it}{N}\right) \right\} \quad (1)$$

Donde a_0, a_i, b_i constantes con $i = 1, \dots, N$. De acuerdo con la ecuación ya es posible analizar los registros y obtener conclusiones acerca del comportamiento de los datos durante el experimento (Kokoszka, Reimherr., 2017).

2.2. Descriptivos

Sea x_1, x_2, \dots, x_n un conjunto de datos funcionales definidos en $t \in T = [a, b] \in \mathbb{R}$ las funciones descriptivas clásicas univariadas y bivariadas están expresadas a continuación (Ramsay and Silverman., 2005).

1. Media: $\hat{x}(t) = \frac{\sum_{i=1}^n x_i(t)}{n}$
2. Varianza: $Var(x(t)) = \frac{\sum_{i=1}^n (x_i(t) - \hat{x}(t))^2}{n-1}$
3. Desviación estándar: $\sqrt{Var(x(t))} = \hat{\sigma}(t)$

4. Covarianza: $Cov(x(t_1), x(t_2)) = \frac{\sum_{i=1}^n (x_1(t_1) - \hat{x}(t_1))(x_1(t_2) - \hat{x}(t_2))}{n-1}$
5. Correlación: $Cor(x(t_1), x(t_2)) = \frac{Cov(x(t_1), x(t_2))}{\sqrt{Var(x(t_1))Var(x(t_2))}}$

2.3. Modelo para datos funcionales dispersos

Sean n acciones independientes en L^2 de un proceso estocástico $(X(t) : t \in [0, 1])$, observadas en una sucesión de puntos en $[0, 1]$ con “ruido de media aditivo”, lo que indica que los datos observados $\{Y_{ij} : 1 \leq j \leq m_i; 1 \leq i \leq n\}$, pueden llegar a ser modelados por:

$$Y_{ij} = X_i(t_{ij}) + \sigma \epsilon_{ij}, \quad (2)$$

donde ϵ_{ij} son i.i.d con media 0 y varianza 1. Teniendo que $X(t)$ proviene de un proceso estocástico en L^2 , por medio del teorema Mercer (*Peng y Paul, 2013*), se establece que existe un kernel semi-definido positivo $C(.,.)$ tal que $Cov(X_s, X_t) = C(s, t)$, y su representación es:

$$X_i(t) = \mu(t) + \sum_{v=1}^{\infty} \sqrt{\lambda_v} \psi_v(t) \xi_{iv} \quad (3)$$

donde la media de la función está definida por $\mu(.) = \mathbb{E}(X(.))$; $\lambda_1 \neq \lambda_2 \geq \dots \geq 0$ son los valores propios del kernel $C(.,.)$; $\psi_v(.)$ son las funciones propias ortogonales y $\{\xi_{iv} : v \geq 1\}$ variables aleatorias para cada i , se determinan no correlacionadas con media 0 y varianza 1.

En la ecuación 2, se asume que $\{t_{ij} : j = 1, \dots, m_i\}$ se muestrea aleatoriamente a partir de una distribución continua. En la mayoría de los casos, m_i suele ser pequeño.

El suavizado de las funciones propias indica que pueden llegar a estar aproximadas a una buena base estable para las clases de funciones suaves, como por ejemplo, la base B - Spline (*Jie y Debashis., 2012*). Ahora, si se asume de igual forma para la segunda ecuación, que $\lambda_v = 0$ para $v > 1$ para algunos $r \geq 1$, se puede obtener un conjunto finito de funciones L^2 linealmente independientes $\{\phi_1(.), \dots, \phi_M(.)\}$ con $M \geq r$, tal que las funciones propias pueden llegar a ser modeladas como:

$$\psi_v(.) = \sum_{k=1}^M b_{kv} \phi_k(.) \quad \text{para cada } v = 1, \dots, r. \quad (4)$$

Por lo tanto, para cada t se tendría lo siguiente:

$$\psi(t) = (\psi_1(t), \dots, \psi_r(t))^T = (\phi_1(t), \dots, \phi_M(t))B, \quad (5)$$

$B = ((b_{kv}))$, una matriz $M \times r$, satisface la restricción:

$$B^T \left(\int \phi(t) \phi(t)^T dt \right) B = \int \psi(t) \psi(t)^T dt = I_r \quad (6)$$

donde $\phi(.) = (\phi_1(.), \dots, \phi_M(.))^T$. Teniendo en cuenta que la matriz $M \times M \int \phi(t) \phi(t)^T dt$ es conocida y no singular, sin pérdida de generalidad, de aquí en adelante se asume que $B^T B = I_r$, por ortogonalidad $\{\phi_1(.), \dots, \phi_M(.)\}$.

Aquí se está asumiendo un modelo de rango reducido para el kernel de la covarianza. Suponiendo que el kernel de la covarianza $C(s, t)$ del modelo tiene una expansión infinita de Karhunen-Loève:

$$C(s, t) = \sum_{k=1}^{\infty} \lambda_k \psi_k(s) \psi_k(t), \quad (7)$$

donde $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, $\sum_{k=1}^{\infty} \lambda_k < \infty$, $y\{\psi_k\}_{k=1}^{\infty}$, forman una base completamente ortogonal para $L^2[0, 1]$. La condición $\sum_{k=1}^{\infty} \lambda_k < \infty$ implica tanto que $\lambda_k \rightarrow 0$, como que $k \rightarrow \infty$. Del mismo modo, la ortogonalidad de las funciones propias $\{\psi_k\}$, implica que ψ_k se vuelva cada vez más volátil cada que k incremente. La manera de truncar la serie de la ecuación de la covarianza del kernel (ecuación 7), para el mismo $r \geq 1$, es:

$$C_{proj}^r(s, t) = \sum_{k=1}^r \lambda_k \psi_k(s) \psi_k(t), \quad (8)$$

lo cual se denomina el kernel de la covarianza proyectado. Notese que $\|C - C_{proj}^r\|_F^2 = \sum_{k=r+1}^{\infty} \lambda_k^2$. Esto, a medida que los valores propios tiendan a 0 rápidamente, incluso con una r pequeña, la aproximación del kernel de la covarianza proyectado, solo resulta con un pequeño sesgo.

Si se asume normalidad en los procesos, $\xi_{iv} \sim^{i.i.d} N(0, 1)$ y $\epsilon_{ij} \sim^{i.i.d} N(0, 1)$ e independencia, bajo el supuesto de $\psi(t) = (\psi_1(t), \dots, \psi_r(t))^T = (\phi_1(t), \dots, \phi_M(t))B$. Se tiene que la log-verosimilitud negativa de los datos, con la condición de $\{(m_i, \{T_{ij}\}_{j=1}^{m_i})\}_{i=1}^n$, está dada por:

$$\begin{aligned} -\log L(B, \Lambda, \sigma^2) &= \frac{1}{2} \sum_{i=1}^n \text{Tr}[(\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i)^{-1} (Y_i - \mu_i)(Y_i - \mu_i)^T] \\ &+ \frac{1}{2} \sum_{i=1}^n \log |\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i|, \end{aligned} \quad (9)$$

donde Λ es la matriz diagonal de tamaño $r \times r$ de valores propios diferentes de cero de $C(\cdot, \cdot)$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^T$, $\mu_i = (\mu(T_{i1}), \dots, \mu(T_{im}))^T$ son los vectores $m_i \times 1$, y $\Phi_i = [\phi(T_{i1}) : \dots : \phi(T_{im})]$ es una matriz de tamaño $M \times n$.

La dificultad de trabajar con la función de verosimilitud reside en la ecuación 9, y que el parámetro B tiene restricciones (ecuación 6). De igual forma, se presenta un problema de optimización no convexa con respecto a los parámetros, puesto que la hessiana de la función objetivo no es globalmente definida positiva. Adicional a lo anterior, el supuesto de normalidad, solo se utiliza con fines de trabajo matemático, ya que la función de la log-verosimilitud negativa es una función de pérdida de *buena fé*. Además, se asume que dicha función es diferenciable con respecto a los valores y funciones propias, lo que a su vez depende de que todos los valores propios diferentes de cero de C sean distintos.

Para el proceso de estimación, es necesario el uso del algoritmo Newton-Raphson. Según Edelman *et.al*(1998), se debe derivar y conjugar el gradiente del algoritmo de Newton-Raphson para optimizar las funciones en los colectores de Stiefel y Grassman (Peng y Paul, 2013). Estos algoritmos tienen como propósito establecer el gradiente de la función objetivo a 0. Para el caso expuesto en este documento, se propone utilizar el algoritmo de Newton-Raphson para hallar el estimador de máxima verosimilitud, debido a su rápida convergencia.

Para comenzar con la minimización de la función de pérdida (log-verosimilitud negativa), se reescribe de la siguiente manera:

$$F(B, \Lambda, \sigma^2) = \sum_{i=1}^n [F_i^1(B, \Lambda, \sigma^2) + F_i^2(B, \Lambda, \sigma^2)]$$

donde

$$F_i^1(B, \Lambda, \sigma^2) = Tr[P_i^{-1} \tilde{Y}_i \tilde{Y}_i^T] \quad y \quad F_i^2(B, \Lambda, \sigma^2) = \log|P_i|$$

con $P_i = \sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i$, $\tilde{Y}_i = Y_i - \mu_i$. Aquí se trata a μ conocida, puesto que se pretende estimar por separado mediante una metodología lineal. Los espacios de parámetros para Λ y σ^2 se definen como conos positivos en espacios euclidianos y, de igual manera, convexos. Por último, el espacio de parámetros para B es $S_{M,r}$, la variedad Stiefel de matrices de tamaño $M \times r$ con columnas ortonormales.

3. Electromiografía

La relación existente entre los componentes de la emoción, con la acción muscular, es estudiada por medio de varias reacciones eléctricas generadas en espacios específicos del cerebro. Existe un método científico encargado del estudio de dichas activaciones eléctricas, denominado Electromiografía (EMG), el cual, por medio de electrodos insertados en los músculos, o a través de electrodos levemente colocados en la superficie de la piel, mide cada paso eléctrico de los músculos en forma de ondas, que a su vez es transformada en una señal, aquella que le indica al especialista la reacción o emoción de un individuo al momento de la exposición a alguna actividad específica.

Formalmente, cuando una unidad motora o motor neurona se activa, libera un impulso denominado potencial de acción, el cual se desplaza de la neurona hacia el músculo, aquel contacto es llamado unión neuromuscular. Luego de que el impulso se transmite a través de la unión neuromuscular, se genera un potencial en todas las fibras musculares provenientes de la unidad motora particular. Todo este proceso eléctrico es conocido como un potencial motor de la acción de la unidad (MUAP). La suma de los potenciales de cada una de las unidades motoras, es aquella señal que se evalúa durante un EMG (M.B.I. Raez, 2006).

Como se expuso anteriormente, el proceso para la evaluación por medio de EMG se puede determinar por medio de dos metodologías, la primera es mediante inyecciones de electrodos insertados a través de la piel hasta el tejido muscular, un profesional observa, al mismo tiempo, la actividad eléctrica mientras se le es implantado el electrodo. Toda información obtenida desde el inicio hasta el final del procedimiento es de gran validez, puesto que el músculo mientras está en reposo genera unas señales normales, pero en cuenta el electrodo es insertado la actividad eléctrica se vuelve anormal, generando las señales necesarias para el análisis correspondiente (Kleissen RF, 1998).

La segunda metodología es por medio de la colocación de electrodos sobre la superficie de la piel, el cual, es utilizado en el presente análisis. Este procedimiento consta de colocar tres (3) electrodos en los músculos especiales del rostro (frente y pómulos). Cada electrodo está encargado de recibir todos los impulsos eléctricos de los músculos, los cuales son generados por las reacciones faciales ante cualquier estímulo. Cada uno de esos impulsos son conducidos por señales, las cuales indican disgusto, preferencias, análisis, entre otras emociones, provocadas por el estímulo expuesto. Cada señal es transformada en una onda, la cual, es leída para la posterior inferencia y análisis.

4. Resultados

La base de datos fue proporcionada por el grupo de investigación Consumo y Mercados del programa académico de Mercadeo de la Universidad Santo Tomás. La investigación consistió en tomar a un grupo

de individuos entre 18 a 35 años de la Universidad Santo Tomás con el propósito de conocer la percepción acerca de los aspectos emocionales frente a la compra de bebidas lácteas saludables en redes sociales.

A cada uno de los sujetos le fueron conectados una serie de electrodos que tenían como función medir la respuesta electromiográfica al nivel de la comisura de los labios, una vez ubicados los electrodos, se le presentaron seis diferentes estímulos referentes a seis diferentes marcas de productos lácteos saludables, para este caso el sujeto no debía realizar ninguna tarea, solo debía observar cada uno de los estímulos mientras se monitoreaba con el electromiógrafo su actividad eléctrica al nivel de la comisura de los labios.

Cada uno de los estímulos fue presentado por un tiempo aproximado de 30 segundos, luego de finalizada esta fase, se le pidió a cada sujeto diligenciar un cuestionario que tenía como fin evaluar sus ideas acerca del consumo de bebidas lácteas funcionales. El cuestionario estaba compuesto por 45 ítems con opción de respuesta tipo escala Likert en un escala de 1 a 6. El cuestionario fue estructurado para evaluar tres aspectos importantes, el primero evalúa la intención de compra de bebidas lácteas saludables, el segundo aspecto evalúa la opinión de cada uno de los sujetos frente a las bebidas lácteas saludables y el tercer aspecto evalúa la acción de compra de bebidas lácteas saludables. En total fueron evaluados 39 sujetos en las dos fases.

La metodología de análisis estadístico propuesta para esta investigación consiste en analizar los primeros 1700 ms después de presentado el estímulo, por medio de análisis de componentes principales funcionales para datos dispersos, debido a que existen registros con diferente longitud de medida de tiempo. A partir del análisis mencionado, se extraerán las coordenadas correspondientes a cada uno de los factores retenidos en el análisis, para luego realizar un análisis de conglomerados por el método de k -medias e identificar los grupos de señales que muestren niveles de agrado y desagrado frente a los estímulos presentados. Por otra parte, se realizará un análisis factorial al cuestionario aplicado a los sujetos, con el propósito de validar por medio del criterio α de cronbach los tres aspectos teóricos planteados e identificar los ítems que cargan en cada uno de los factores, para luego identificar los individuos que son representativos para cada uno de los factores.

Finalmente, luego de identificar los sujetos tanto el análisis de conglomerados como en el análisis factorial, se construirá una tabla de contingencias en la que se puede identificar cómo se agrupan los sujetos. A partir de esto, se realizará un análisis de correspondencias múltiples para encontrar la relación entre la evaluación hecha por los sujetos en el cuestionario y la señal electromiográfica obtenida por los estímulos presentados.

5. Análisis Factorial

En esta etapa, se analizarán las respuestas al cuestionario aplicado a cada uno de los sujetos que hicieron parte del experimento; de acuerdo con la estructura del cuestionario y de los datos, se aplicará un Análisis Factorial con el propósito de validar cada uno de los ítems dentro de los factores teóricos construidos. Por medio del criterio α de Cronbach se buscará el mejor conjunto de ítems que se ajusten a cada uno de los factores, de modo que solo serán tenidos en cuenta los ítems que presenten una ponderación mayor 0.5 para ser clasificados dentro del factor latente.

La batería de ítems fue estructurada en tres aspectos, (cada uno compuesto por 15 ítems) que se consideraron de importancia para evaluar el comportamiento de consumo de bebidas lácteas saludables en los individuos, estos tres aspectos son: Intención de compra, Opinión y Acción de compra.

Por medio de la prueba de Bartlett se evidencio que la matriz de correlación es diferente de la matriz identidad, con un valor Chi-cuadrado de 6508.267 y un p-valor de 0, se rechaza la hipótesis nula en favor de la alternativa, indicando que la matriz de correlación es estadísticamente diferente de cero y por lo tanto es adecuado realizar el análisis factorial.

5.1. Validación ítems para cada aspecto

5.1.1. Validación ítems del aspecto intención de compra

En la intención de compra, se midió el interés que había por parte de los individuos que participaron en el experimento a consumir bebidas lácteas saludables, de acuerdo con lo anterior, se muestra cada uno de los ítems evaluados.

- P1. Mi preocupación por mi cuerpo y mi salud es menor comparado con los malestares que padezco.
- P2. La mala nutrición se debe a la desinformación por parte de los medios de comunicación.
- P3. Aquellas personas que se preocupan por su salud deben ser apreciadas y respetadas por lo que hacen.
- P4. Me enoja ver la imagen de un famoso en un anuncio publicitario consumiendo un yogurt saludable.
- P5. Compraría alguna bebida láctea saludable si supiera que con esto contribuye a mi bienestar físico y emocional.
- P6. Compraría alguna bebida láctea saludable si está respaldada por una sociedad médica.
- P7. Compraría alguna bebida láctea saludable si posee una etiqueta o sello de calidad alimentaria.
- P8. Compraría alguna bebida láctea saludable si posee una etiqueta o sello medio ambiental.
- P9. Los seres humanos tienen derecho a modificar sus hábitos alimenticios en función de sus necesidades y deseos.
- P10. Compraría alguna bebida láctea saludable aunque fuera mucho más costosa que las bebidas lácteas tradicionales.
- P11. Estoy dispuesto a realizar diferentes actividades para promover los hábitos y estilos de vida saludables.
- P12. Compraría bebidas lácteas saludables si éstas fueran promocionadas a través de las redes sociales.
- P13. Los seres humanos están abusando en exceso de las redes sociales.
- P14. Con seguridad la próxima semana compraré bebidas lácteas saludables.
- P15. Estoy dispuesto a comprar una bebida láctea saludable basándome en el etiquetado nutricional.

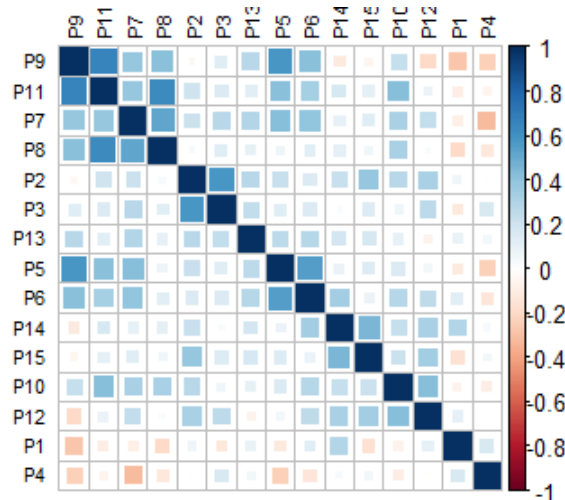


Figura 1: Matriz de correlación de Pearson del factor intención de compra

En la figura 1 se observa la matriz de correlación de Pearson para los 15 ítems que conforman el aspecto teórico intención de compra.

Luego de aplicar el análisis factorial ajustado con un factor para el aspecto de intención de compra tabla 1, se observa que al retener los ítems con una ponderación mayor a 0.5 y realizando un nuevo análisis factorial se logra obtener un alfa de Cronbach de 0.814 superior al obtenido con todos los ítems que fue de 0.743, es decir, que estos ítems logran conformar el factor del aspecto intención de compra con una mayor fiabilidad.

Tabla 1: Validación de ítems del factor intención

	Item	Cargas	Cargas.1
1	P1	-0.16	
2	P2	0.28	
3	P3	0.28	
4	P4	-0.23	
5	P5	0.62	0.62
6	P6	0.56	0.51
7	P7	0.65	0.57
8	P8	0.58	0.58
9	P9	0.72	0.81
10	P10	0.48	
11	P11	0.77	0.80
12	P12	0.17	
13	P13	0.34	
14	P14	0.20	
15	P15	0.20	
16	α Cronbach	0.74	0.81

5.1.2. Validación ítems del aspecto opinión

En este aspecto se evaluó la opinión de cada uno de los individuos con relación a la publicidad, respaldo médico y demás temas que se consideraron relevantes con respecto al consumo de bebidas lácteas saludables. Los siguientes ítems son aquellos que conforman el factor opinión sobre las bebidas lácteas saludables, con su respectiva gráfica de correlación de Pearson para estos ítems en la figura 2.

- P16. Conocer los efectos en la salud de las bebidas lácteas que consume.
- P17. Reconocer los componentes y propiedades que contienen las bebidas lácteas saludables que consume.
- P18. Tener acceso a información nutricional por parte de las marcas de bebidas lácteas saludables.
- P19. Que existan organismos internacionales que propendan por controlar la publicidad engañosa en torno a los alimentos saludables.
- P20. Influir en otras personas para que consuman productos lácteos saludables.
- P21. Que universidades, gobierno y otras organizaciones investiguen sobre la alimentación saludable y el consumo responsable.

- P22. Consumir alimentos o productos saludables aunque su precio sea mayor.
- P23. Apoyar el ejercicio y la sana alimentación.
- P24. Comprar bebidas lácteas saludables en lugares certificados.
- P25. Que el gobierno haga leyes para que se regule la información en redes sociales.
- P26. Pertenecer a grupos promotores de alimentación y estilos de vida saludable.
- P27. Que las empresas respeten las normas legales en cuanto a la comunicación publicitaria de la información alimentaria.
- P28. Averiguar qué tan saludable es una bebida láctea.
- P29. Que los anuncios publicitarios transmitan información más confiable y real.
- P30. Que las marcas de bebidas lácteas saludables manifiesten el respeto por la sociedad y el medio ambiente.

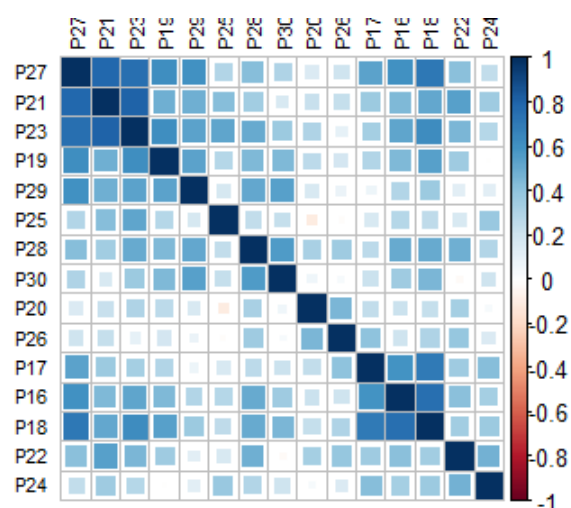


Figura 2: Matriz de correlación de Pearson del factor opinión de compra

Tabla 2: Validación de ítems del factor opinión

De igual forma como se observa en la tabla 2, al aplicar el análisis factorial reteniendo los ítems con una carga mayor a 0.5, se retuvieron 9 ítems con alfa de Cronbach significativamente mayor al inicial, y al realizar una nueva depuración se eliminó el ítem P22 logando así obtener un aumento en el α de Cronbach de 0.906, que sugiere que al retener 8 ítems se logra una mayor fiabilidad dentro del factor latente opinión.

5.1.3. Validación ítems del factor acción de compra

Por último, el aspecto acción de compra evaluó la intención real de cada uno de los individuos de comprar una bebida láctea saludable. A continuación, se muestran los ítems que componen el factor de acción de compra.

	Item	Cargas	Cargas.1	Cargas.2
1	P16	0.70	0.67	0.67
2	P17	0.57		
3	P18	0.80	0.76	0.77
4	P19	0.68	0.69	0.69
5	P20	0.30		
6	P21	0.80	0.83	0.82
7	P22	0.53	0.52	
8	P23	0.86	0.87	0.86
9	P24	0.38		
10	P25	0.42		
11	P26	0.29		
12	P27	0.88	0.90	0.91
13	P28	0.60	0.57	0.56
14	P29	0.59	0.62	0.63
15	P30	0.44		
16	α Cronbach	0.74	0.91	0.91

- P31. Consumir alimentos nutritivos, fortificados y bajos en grasa.
- P32. Examinar los componentes de la etiqueta nutricional de la bebida láctea antes de comprarla.
- P33. Seguir en redes sociales las promociones de las marcas de bebidas lácteas para luego comprar.
- P34. Comprar bebidas lácteas saludables que tengan respaldo de una sociedad médica.
- P35. Comprar la bebida láctea por las bondades del anuncio publicitario en las redes sociales.
- P36. Comprar y consumir la bebida láctea saludable según el problema de salud que padezco.
- P37. Comprar la bebida láctea saludable para parecerme a la imagen fitness del anuncio publicitario.
- P38. Hacer ejercicio y llevar una dieta sana y balanceada.
- P39. Consumir alimentos altos en grasa y en concentración calórica.
- P40. Sentir culpa por haber consumido un alimento alto en calorías después de ver un anuncio en redes sociales.
- P41. Consumir bebidas lácteas saludables a las medias nueves u onces.
- P42. Pertenecer a grupos en redes sociales con una finalidad saludable.
- P43. Comprar las bebidas lácteas saludables a partir de los comentarios y recomendaciones en las redes sociales.
- P44. Comprar una bebida láctea saludable por el número de me gustas en las redes sociales.
- P45. Comprar una marca de bebida láctea saludable en función de su contribución al medio ambiente y social.

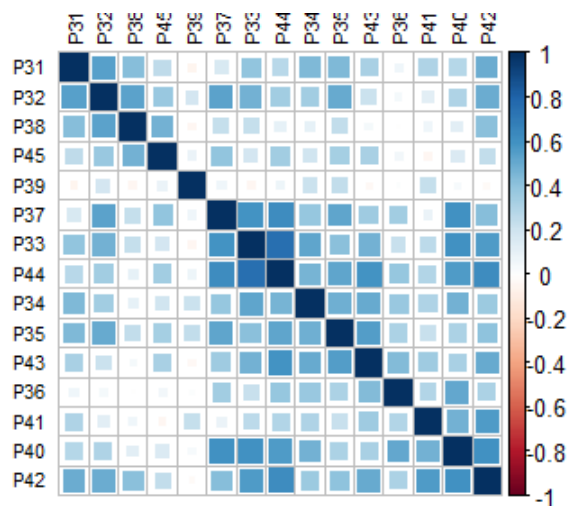


Figura 3: Matriz de correlación de Pearson del factor acción de compra

Tabla 3: Validación de ítems del factor acción

	Item	Cargas	Cargas.1
1	P31	0.50	0.48
2	P32	0.58	0.57
3	P33	0.80	0.83
4	P34	0.63	0.62
5	P35	0.65	0.64
6	P36	0.44	
7	P37	0.71	0.71
8	P38	0.32	
9	P39	0.26	
10	P40	0.71	0.70
11	P41	0.43	
12	P42	0.75	0.73
13	P43	0.64	0.63
14	P44	0.83	0.84
15	P45	0.39	
16	α Cronbach	0.88	0.90

Finalmente luego de replicar los mismos pasos que se hicieron en los dos primeros factores, en la tabla 3 se observa que al retener 10 factores se logra obtener una mayor fiabilidad de la escala dentro del factor con un α de Cronbach de 0.896.

5.2. Análisis factorial luego de la validación

Una vez hecha la validación del instrumento en cada uno de los aspectos teóricos planteados, se procede a construir el análisis factorial completo y a partir de este obtener las cargas de individuos en los factores, que serán de utilidad para la tercera parte del análisis.

Tabla 5: Varianzas del Análisis factorial completo

	Acción	Opinión	Intención
SS loadings	5.21	4.98	3.19
Proportion Var	0.22	0.21	0.13
Cumulative Var	0.22	0.42	0.56

Tabla 4: Análisis factorial completo

	Ítems	Acción	Opinión	Intención
1	P5			0.43
2	P6			0.43
3	P7			0.48
4	P8			0.69
5	P9			0.72
6	P11			0.91
7	P16		0.61	
8	P18		0.73	
9	P19		0.67	
10	P21		0.83	
11	P23		0.85	
12	P27		0.92	
13	P28		0.52	
14	P29		0.64	
15	P31	0.55		
16	P32	0.60		
17	P33	0.83		
18	P34	0.69		
19	P35	0.66		
20	P37	0.68		
21	P40	0.71		
22	P42	0.71		
23	P43	0.61		
24	P44	0.80		

En las tablas 4 y 5 se tiene el resultado obtenido luego de aplicar el método de análisis factorial, en el que, con los 3 factores planteados se retiene el 55.8% de la varianza total, se aplicó el método de rotación Varimax con el propósito de maximizar la varianza en las componentes.

En la tabla 6 se puede evidenciar cada uno de los factores con los respectivos ítems con los cuales

quedaron compuestos cada uno de los aspectos teóricos que se plantearon en el cuestionario.

A partir del resultado obtenido en este análisis, ya es posible obtener las cargas de los individuos en los factores y poder continuar con el análisis del proyecto.

Tabla 6: Ítems válidos

Aspecto	Ítems
Intención	<p>P5. Compraría alguna bebida láctea saludable si supiera que con esto contribuye a mi bienestar físico y emocional</p> <p>P6. Compraría alguna bebida láctea saludable si está respaldada por una sociedad médica</p> <p>P7. Compraría alguna bebida láctea saludable si posee una etiqueta o sello de calidad alimentaria</p> <p>P8. Compraría alguna bebida láctea saludable si posee una etiqueta o sello medio ambiental</p> <p>P9. Los seres humanos tienen derecho a modificar sus hábitos alimenticios en función de sus necesidades y deseos</p> <p>P11. Estoy dispuesto a realizar diferentes actividades para promover los hábitos y estilos de vida saludables.</p>
Opinión	<p>P16. Conocer los efectos en la salud de las bebidas lácteas que consume</p> <p>P18. Tener acceso a información nutricional por parte de las marcas de bebidas lácteas saludables.</p> <p>P19. Que existan organismos internacionales que propendan por controlar la publicidad engañosa en torno a los alimentos saludables</p> <p>P21. Que universidades, gobierno y otras organizaciones investiguen sobre la alimentación saludable y el consumo responsable</p> <p>P22. Consumir alimentos o productos saludables aunque su precio sea mayor</p> <p>P23. Apoyar el ejercicio y la sana alimentación</p> <p>P27. Que las empresas respeten las normas legales en cuanto a la comunicación publicitaria de la información alimentaria</p> <p>P28. Averiguar qué tan saludable es una bebida láctea</p> <p>P29. Que los anuncios publicitarios transmitan información más confiable y real</p>
Acción	<p>P31. Consumir alimentos nutritivos, fortificados y bajos en grasa</p> <p>P32. Examinar los componentes de la etiqueta nutricional de la bebida láctea antes de comprarla</p> <p>P33. Seguir en redes sociales las promociones de las marcas de bebidas lácteas para luego comprar</p> <p>P34. Comprar bebidas lácteas saludables que tengan respaldo de una sociedad médica</p> <p>P35. Comprar la bebida láctea por las bondades del anuncio publicitario en las redes sociales</p> <p>P37. Comprar la bebida láctea saludable para parecerme a la imagen fitness del anuncio publicitario</p> <p>P38. Hacer ejercicio y llevar una dieta sana y balanceada</p> <p>P39. Consumir alimentos altos en grasa y en concentración calórica</p> <p>P40. Sentir culpa por haber consumido un alimento alto en calorías después de ver un anuncio en redes sociales</p> <p>P42. Pertenecer a grupos en redes sociales con una finalidad saludable</p> <p>P43. Comprar las bebidas lácteas saludables a partir de los comentarios y recomendaciones en las redes sociales</p> <p>P44. Comprar una bebida láctea saludable por el número de me gusta en las redes sociales</p>

6. Análisis de datos funcionales EMG

6.1. Descriptivos

A continuación se presentan algunos descriptivos del comportamiento de la señal de cada uno de los individuos para cada estímulo presentado.

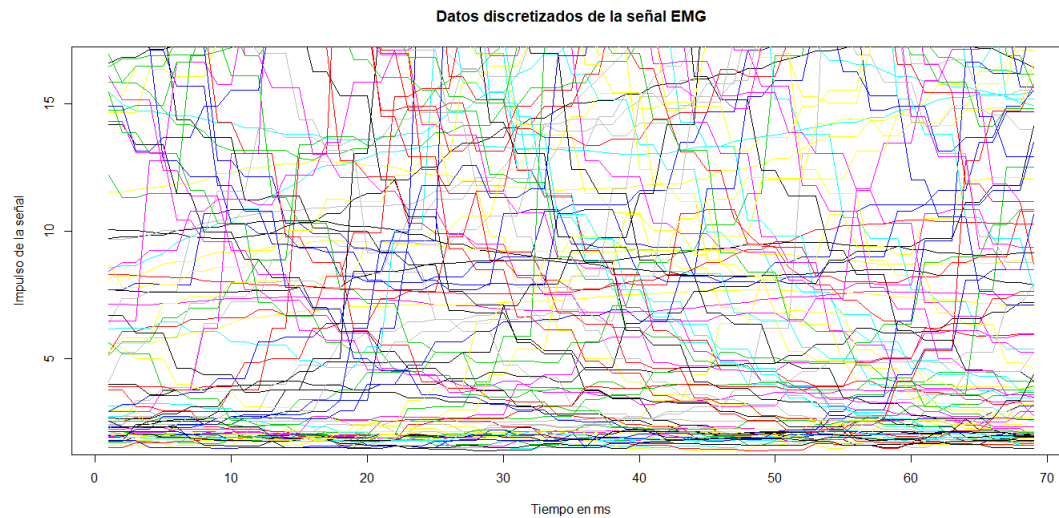


Figura 4: Datos discretizados

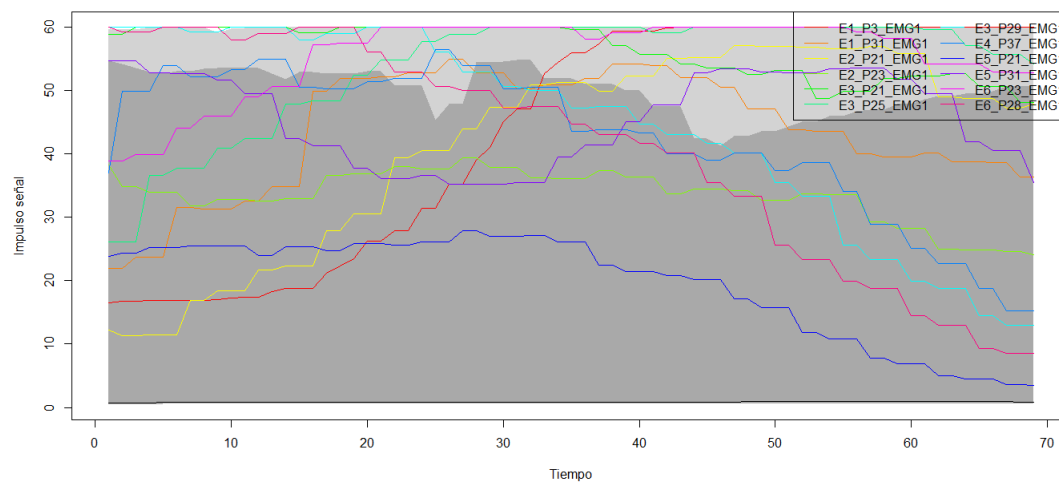


Figura 5: Funciones atípicas

La figura 4 muestra el comportamiento de la señal capturada por el electrodo ubicado al nivel de la comisura de los labios para cada individuo luego de presentarle el estímulo, cada señal representa 69 registros medidos cada 24 o 25 ms, para un registro total de aproximadamente 1.700 ms, dependiendo del individuo, esto debido a que el equipo en algunos casos tomaba el registro en diferentes rangos de tiempo.

La figura 5 muestra el gráfico de bagplot funcional (Hyndman., 2010) para datos funciones, en este se puede observar que existen 12 señales que presentan un comportamiento atípico en relación con la demás señales capturadas.

En este punto los registros se están discretizados y por lo tanto no es posible trabajar con los datos de forma adecuada, es por eso que es necesario realizar una transformación de los datos para convertirlos en funciones por medio de bases de Fourier.

6.2. Suavizados por serie de Fourier

Por medio del software estadístico R y usando la librería *fda*, se analizaran los registros para transformar los datos en funciones por medio de una aproximación de series de Fourier como se muestra a continuación.

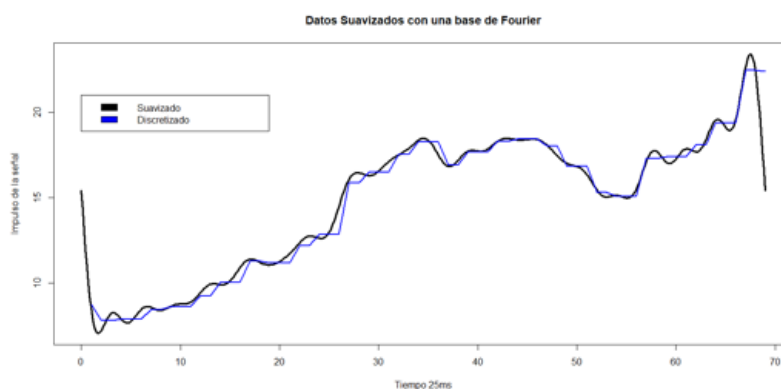


Figura 6: Aproximación por series de Fourier

En la figura 6 se puede observar el comportamiento de la señal EMG discretizada y luego de convertirla en una función por medio de una aproximación a series de Fourier, se evidencia que conserva la misma relación en su comportamiento, en este punto ya es posible trabajar con los registros como datos funcionales. La figura 7 muestra las funciones EMG luego de ser suavizadas.

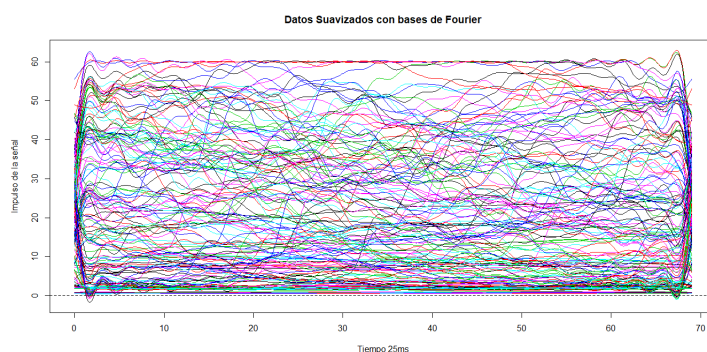


Figura 7: Suavizado por series de Fourier

6.3. Función media, desviación estándar y correlación

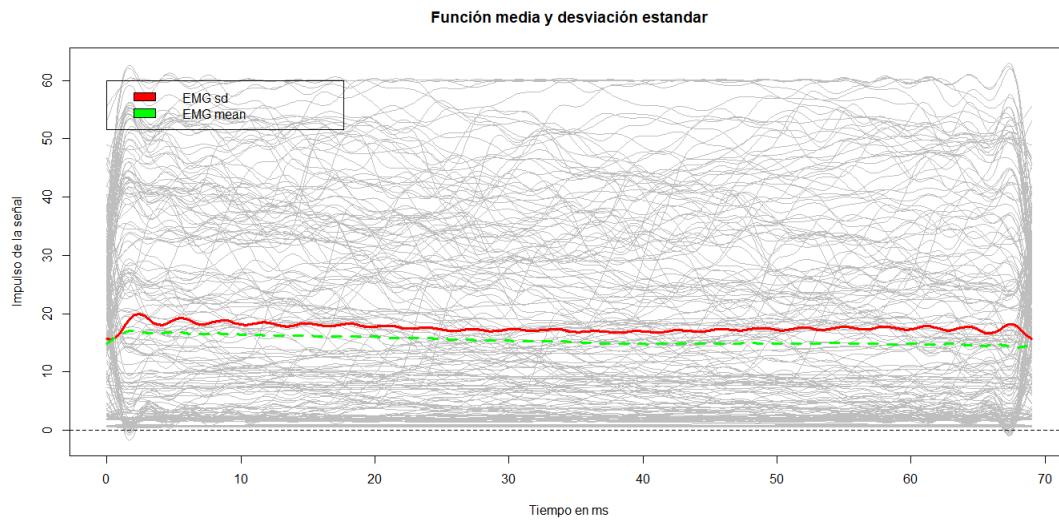


Figura 8: Función media y desviación estándar

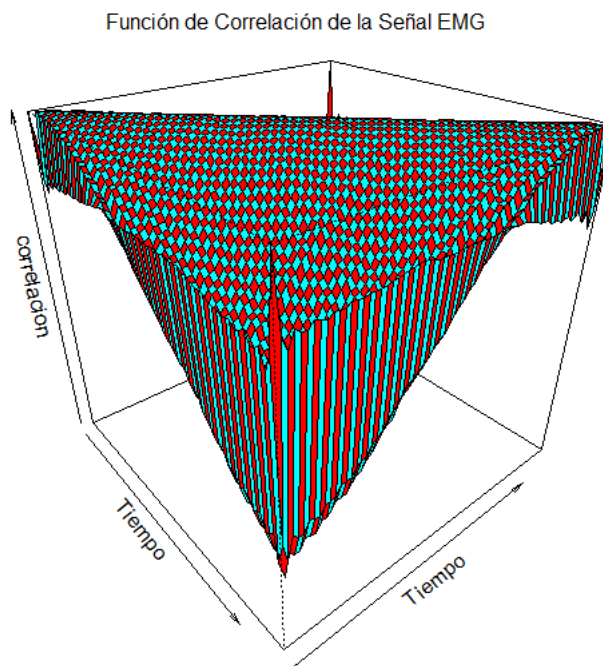


Figura 9: Función de correlación

Se observa un comportamiento bastante plano en cuanto a la media y la desviación estándar de las ondas, con valores alrededor 18 puntos, que es el impulso de la señal EMG generada por los estímulos, si se analiza detalladamente este resultado, se pensaría que hay un gran grupo de ondas que se encuentran con un comportamiento muy plano por debajo de los 20 puntos en el impulso de la señal, lo que daría un indicio para pensar que en este grupo de ondas no hubo una reacción fuerte por parte los estímulos presentados.

6.4. Análisis de Componentes Principales Funcionales para datos dispersos (ACPF-D)

Por medio del análisis de componentes principales para datos dispersos, y haciendo uso de la librería *fpca*, se modelarán las funciones EMG, esto debido a lo que se mencionó en la parte descriptiva de los datos. Existen dos grupos de individuos en el que sus registros fueron marcados cada 24 ms y otro grupo sus registros fueron marcados cada 25 ms, debido a esta diferencia de tiempos en los registros, se hace necesario aplicar una metodología diferente a la conocida para modelar datos funcionales con el método de ACP:

La metodología de ACPF-D tiene como finalidad emparejar esta diferencia en los tiempos y por medio de máxima verosimilitud restringida por el método de Newton-Raphson estimar la autocovarianza de las funciones, para luego estimar los valores y funciones propias y de esta forma poder describir los modos de variación de las ondas EMG.

Vemos que la curva media estimada por el método de ACPF-D (figura 10) se aproxima muy bien a la curva media calculada en la parte descriptiva de esta sección.

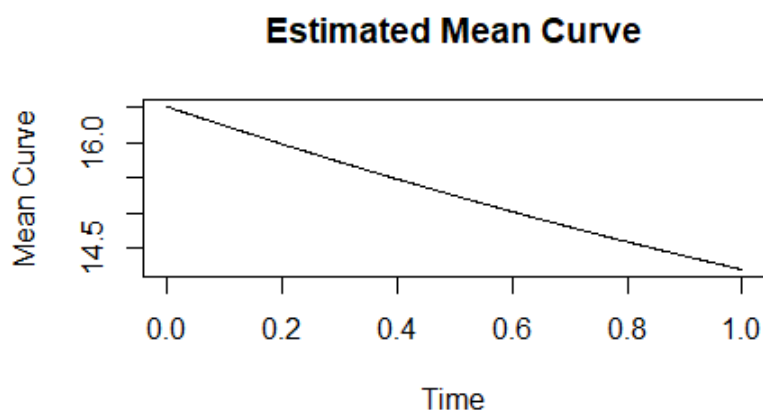


Figura 10: Curva media estimada por el método ACPF-D

Tabla 7: Resultados del ACPF-D

	Componente 1	Componente 2
Valor propio	241,495	56,438
Proporción	78,7	18,4
Proporción acumulada	78,7	97,1

En la tabla 7 se tienen los resultados obtenidos luego de aplicar el método de ACPF-D, se concluye que al retener dos componentes se recoge el 97.1% de la variación de las ondas EMG, el siguiente paso será

obtener las coordenadas de las componentes retenidas para luego hacer un análisis de conglomerados por método de k -medias funcionales y poder clasificar las ondas EMG de acuerdo con su característica.

6.5. Conglomerados por el método de k -medias

Luego de obtener las coordenadas en el ACPF-D, el siguiente paso es construir el conjunto de conglomerados por el método de k -medias que nos permita clasificar las ondas EMG de acuerdo con su característica.

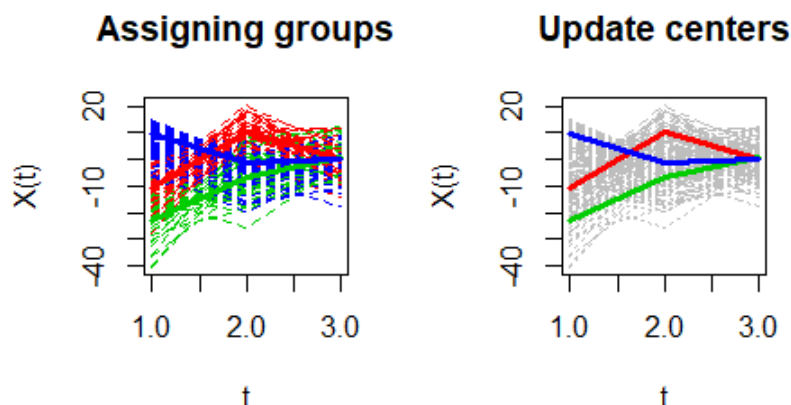


Figura 11: Clasificación cluster de las ondas EMG

Finalmente en la figura 11 se tiene la clasificación cluster de las ondas EMG generados por cada uno de los individuos para cada estímulo. De acuerdo con las características de los centroides funcionales encontrados, las ondas EMG se clasificaron en tres grupos.

El primer cluster fue clasificado por centroide rojo, que de acuerdo con el comportamiento que tiene, son aquellos individuos que al momento de presentarle el estímulo transcurrieron algunos milisegundos antes de comenzar a generar una activación del electrodo ubicado cerca de la comisura de los labios. A este cluster se le dio el nombre Analista.

El segundo cluster, fue clasificado por el centroide verde, este grupo de ondas EMG que se clasificaron en este cluster, de acuerdo con la característica que presenta, son aquellos individuos que al momento de presentarle el estímulo generó una baja activación del electrodo durante la mayor parte que estuvo expuesto al estímulo, lo que se podría interpretar como indiferencia al estímulo presentado. A este cluster se le dio el nombre de Indiferente.

El tercer cluster, fue clasificado por el centroide azul, en este grupo las ondas EMG tienen la característica de generar una alta activación del electrodo justo en el momento de ser presentado el estímulo y a medida que transcurre el tiempo, la actividad comienza a disminuir, esto se puede interpretar que justo en el momento que fue presentado el estímulo generó una reacción en el individuo. A este cluster se le dio el nombre de Activo .

A continuación en la tabla 8 se muestra cómo se clasificaron los individuos de acuerdo con los clusters dentro de cada uno de los estímulos.

Tabla 8: Clasificación de los cluster por estímulo

	Estímulo 1	Estímulo 2	Estímulo 3	Estímulo 4	Estímulo 5	Estímulo 6
Analista	0	1	4	14	14	14
Indiferente	1	1	9	11	7	6
Activo	38	37	26	14	18	19

Frente a estos hallazgos, el siguiente paso será identificar a cada individuo al cluster que pertenece de acuerdo con el estímulo, para finalmente aplicar el método de análisis de correspondencias múltiples.

7. Análisis de Correspondencias Múltiples (ACM)

En esta etapa del análisis, el principal propósito será encontrar una relación entre la calificación que el individuo le dio a cada uno de los ítems dentro de los tres aspectos teóricos construidos y luego validados en el análisis factorial y el comportamiento de las ondas EMG clasifica en los tres cluster por el método de K-Medias para datos funcionales.

Para este análisis, la base se estructuró identificando a cada individuo dentro de los tres factores construidos en el análisis factorial de acuerdo con las cargas factoriales y de igual forma se identificó a los individuos dentro los clusters generados por el método de k -medias para datos funcionales de acuerdo con las ondas EMG para cada estímulo.

De acuerdo con los aspectos calificados en el cuestionario, el factor 1 del ACM se caracteriza en su parte negativa por estar representado por individuos que tienen la intención de comprar bebidas lácteas saludables y en su parte positiva está caracterizado por la opinión generada frente a las bebidas lácteas saludables. El factor 2 se caracteriza en su parte positiva por representar individuos que tienen la acción o la intención real de consumir productos lácteos saludables.

En el gráfico de ACM (figura 12), se observa que los estímulos 4, 5, 6 están caracterizados por haber generado una reacción activa en los individuos y están relacionados con la acción o intención real de comprar una bebida láctea saludable, en cuanto al estímulo 3 y 6 que se caracterizan por ser indiferente y analista respectivamente, está relacionado con la intención de compra, finalmente los estímulos 1 y 2 que están caracterizados por ser activos, se encuentran en un punto medio entre la acción, la intención y la opinión de comprar bebidas lácteas saludables, este último se podría interpretar como existencia de una división por parte de los individuos en cuando a cada uno los aspectos calificados por los individuos.

En términos generales, individuos que se ubiquen en el cuadrante 1 se caracterizan por ser indiferentes frente a los estímulos presentados, pero sus puntuaciones en los ítems de opinión se encuentran por arriba del promedio, lo que indicaría que son personas que no tienen ni la intención ni la acción de comprar bebidas lácteas saludables, pero son activos al momento de opinar acerca de las mismas. Por otro lado, personas que se ubiquen en el segundo cuadrante de la gráfica, se caracterizan por ser muy activos y analistas frente a los estímulos generados y por ende son aquellas persona que tienen la intención real de consumir bebidas lácteas saludables.

Individuos que se ubiquen en el tercer cuadrante del gráfico, se caracterizan por ser indiferentes frente a los estímulos presentados, pero muestran algún tipo de intención de consumir o adquirir bebidas lácteas saludables, finalmente, personas que estén ubicados en el cuarto cuadrante, son aquellas personas que se caracterizan por ser muy analistas frente a los estímulos presentados, mostrando algún tipo de intención de compra, pero que frente a estos muestran altos niveles de opinión.

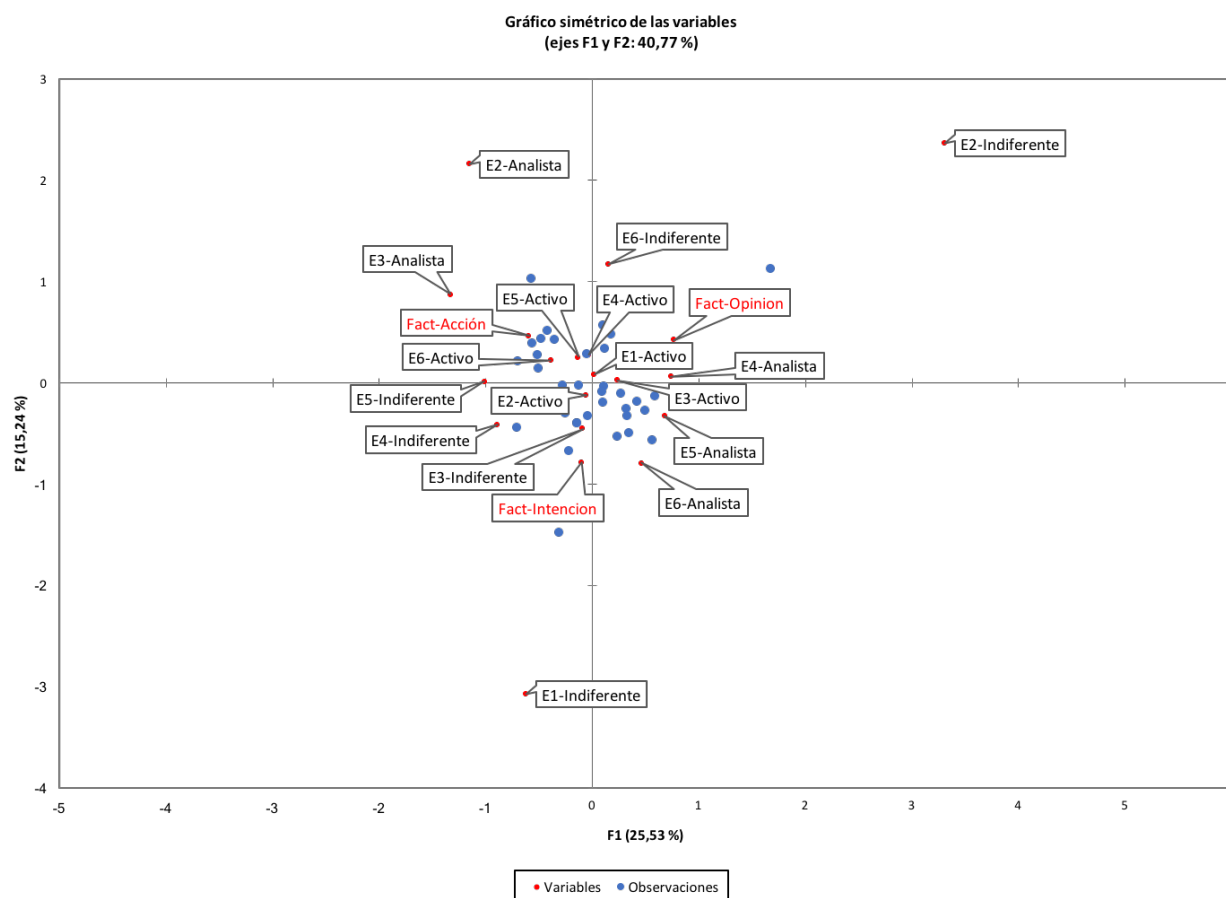


Figura 12: Análisis de correspondencias múltiples

Discusión

A partir de los resultados obtenidos, se pudo evidenciar que por medio del análisis de componentes principales funcionales y en combinación con el análisis de conglomerados para datos funcionales, es posible estructurar una clasificación cluster a partir de las estructuras de las ondas EMG obtenidas en la investigación, permitiendo caracterizar el comportamiento de las estructuras de estas ondas, haciendo posible la interpretación de todas las posibles reacciones que puede tener un individuo frente a los estímulos presentados.

De igual forma, el aplicar un cuestionario a los individuos, permite complementar el análisis en aras de poder dar una interpretación coherente a los resultados obtenido en el análisis de datos funcionales, dado que si no se tiene la experiencia para interpretar la estructura de una onda EMG, puede ser bastante difícil llegar a una conclusión que permita identificar cuál fue la reacción que pudo tener un individuo frente a un determinado estímulo.

Referencias

- [1] Aristizabal, Jeimy. (2011). Metodología estadística para el análisis de datos funcionales cerebrales: Una aproximación con potenciales evocados (Tesis de maestría). *Universidad Nacional de Colombia*, Bogotá D.C., Colombia.
- [2] Camargo, Andrés; Urrea, Carlos; Mogui, José. (2016). Medición de Patrones de Actividad Cerebral Registrados con el Equipo de EEG Emotiv-Epoc, como Indicador de Engaño en Estudiantes Universitarios (Tesis de especialización). *Fundación Universitaria Los Libertadores*, Bogotá D.C., Colombia.
- [3] Carrillo, Alexis; Garatejo, Olga; Pineda, Wilmer. (2017). Análisis multivariado de datos funcionales aplicado a curvas de encefalogramas. *Comunicaciones en Estadística*, 10:129–144.
- [4] Fang Yao; Hans-Georg Müller; Jane-Ling Wang. (2012). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*. 87:470,577–590.
- [5] Gareth M. James; Trevor J. Hastie; Catherine A. Sugar. (2000). Principal Component Models for Sparse Functional Data. *Biometrika Trust*. 87:587–602.
- [6] Hasenstab, Kyle; Scheffler, Aaron; Telesca, Donatello; Sugar, Catherine A.; Jeste, Shafali; Distefano, Charlotte; Sentürk, Damla. (2017). A multi-dimensional functional principal components analysis of EEG data. *Biometrics*.
- [7] Horváth, Lajos; Kokoszka, Piotr. (2012). Inference for Functional Data with Applications. *Springer*
- [8] Jie Peng; Debashis Paul. (2012). A Geometric Approach to Maximum Likelihood Estimation of the Functional Principal Components From Sparse Longitudinal Data. *Journal of Computational and Graphical Statistics*. 18:4, 995–1015.
- [9] Jacques, Julien; Preda, Cristian. (2014). Functional data clustering: a survey. *Springer Verlag*
- [10] Kokoszka, Piotr; Reimherr, Matthew. (2017). Introduction to Functional Data Analysis. *CHAPMAN & HALL/CRC*
- [11] M. B. I. Reaz; M. S. Hussain; F. Mohd-Yasin. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, 8:11–35.
- [12] Ramsay, J.O.; Silverman, B.W. . (2005). Functional Data Analysis. *Springer*
- [13] R.F.M. Kleissen; J.H. Buurke; J. Harlaar; G. Zilvold. (1998). Electromyography in the biomechanical analysis of human movement and its clinical application. *Gait & Posture*, 2:0–158.