

## **Análisis de datos sensoriales de tomate triturado con lógica difusa y técnicas multivariadas**

### **Sensory data analysis crushed tomatoes with fuzzy logic and multivariate techniques**

Matilde Inés Césari<sup>1</sup>, Nancy Beatriz Ventrera<sup>2</sup>, Adriana Gámbaro<sup>3</sup>

Originales: *Recepción*: 04/04/2016 - *Aceptación*: 27/04/2017

#### **RESUMEN**

El conjunto de las técnicas de recolección y análisis de datos, generados en la evaluación sensorial, constituye una rama de la estadística llamada sensometría. Si se tiene en cuenta la vaguedad e incertidumbre con la que se manifiestan las percepciones humanas, la estrategia más apropiada para obtener conocimiento preciso a partir de valoraciones subjetivas, consiste en transformar los datos sensoriales mediante la matemática borrosa. No obstante el potencial que muestra el razonamiento borroso, en la actualidad, existen pocas referencias sobre la aplicación de la lógica difusa en la evaluación sensorial, y no se dispone de criterios y métodos armonizados y sistematizados para una evaluación sensorial objetiva y reproducible, en alimentos vegetales. Los resultados obtenidos en pruebas sensoriales en tomate triturado, permitieron exponer un marco metodológico para el análisis sensométrico de datos borrosos con la utilización de la lógica difusa y métodos de análisis multivariado de datos. De esta manera se optimizan las técnicas en las pruebas sensoriales, para así renovar el proceso de análisis sensorial, desarrollando un área específica que mejorará el valor económico en la producción alimentaria argentina.

#### **Palabras clave**

evaluación sensorial • tomate • datos sensoriales • sensometría • lógica borrosa • análisis de correspondencias

- 
- 1 CeReCon (Centro de Investigación en Computación y Neurociencia). Facultad Regional Mendoza. Universidad Tecnológica Nacional Rodriguez 273. 5500. Mendoza, Argentina. matilde.cesari@yahoo.com  
matilde.cesari@frm.utn.edu.ar
  - 2 Universidad Nacional de Cuyo. Facultad de Ciencias Agrarias. Dpto. de Ciencias Enológicas y Agroalimentarias. Cátedra de Bromatología. Alte. Brown 500. Chacras de Coria. Mendoza. Argentina. M5528AHB. nventrera@fca.uncu.edu.ar
  - 3 Universidad de la República. Facultad de Química. Cátedra de Ciencia y Tecnología de Alimentos. Sección Evaluación Sensorial. Avda. Gral. Flores 2124. C.P.11800. Montevideo. Uruguay. agambaro@fq.edu.uy

## ABSTRACT

All data collection and analysis techniques, generated in the sensory evaluation, constitute a branch of the statistic called sensometry. If it is taking into account the vagueness and uncertainty with which human perceptions are manifested, the most appropriate strategy for obtaining precise knowledge from subjective evaluations is to transform the sensory data through fuzzy mathematics. However the potential of fuzzy reasoning, there are currently few references about the application of fuzzy logic in sensory evaluation, and there are no harmonized and systematized criteria and methods for objective sensory evaluation and reproducibly in plant foods. The results obtained in sensory tests of crushed tomato allowed to expose a methodological framework for the sensometric analysis of fuzzy data with the use of fuzzy logic and methods of multivariate analysis of data. In this way the techniques in the sensorial tests are optimized, and thus the process of sensorial analysis is renewed, developing a specific area that will improve the economic value in the Argentinean food production.

### Keywords

sensory evaluation • tomato • sensory data • sensometrics • fuzzy logic  
• correlation analysis

## INTRODUCCIÓN

El término "datos sensoriales", se aplica a todos aquellos datos generados en la percepción de características de un producto alimenticio, a través de los sentidos. El conjunto de las técnicas de recolección de este tipo de datos y métodos estadísticos de análisis asociados constituyen la rama de la Estadística llamada Sensometría (3), la cual resulta, de gran utilidad en estudios que involucran degustaciones en la industria alimentaria. Estas técnicas se encuentran en continua expansión, dada la creciente competitividad de los mercados, y que genera continuamente nuevos problemas, exigentes de técnicas de análisis de datos más sofisticadas.

Con la finalidad de disponer de medidas sensoriales, de aplicación en el área de productos de origen vegetal, se han adoptado metodologías basadas en el concepto de perfil descriptivo, además de diseño de producto, calidad poscosecha,

articulación de datos sensoriales e instrumentales, y evaluación individual o grupal del rendimiento de un panel de jueces (12, 13, 25, 27).

Una problemática de los métodos de evaluación sensorial (ES), reside en el trabajo con paneles de laboratorio constituidos por personas que proporcionan la información percibida por sus sentidos (16, 22, 28) y se debe lograr que la respuesta humana sea precisa y se pueda determinar de modo cuantitativo.

Existen diferencias en las percepciones sensoriales individuales en cuanto a gustos, sonidos, formas, iluminación, olores, etc., de tal forma que cada sujeto tiene una imagen global del mundo. Esta variabilidad esta reforzada por la educación, nivel social, cultura y personalidad del catador (2), por lo que los resultados de las pruebas sensoriales se expresan en términos relativos y no en términos absolutos, por ende, este tipo de información implica siempre subjetividad o

vaguedad. La incerteza es aquella que surge de procesos humanos, como la sensación, la percepción, la experiencia cognoscitiva, el razonamiento y el pensamiento.

La percepción que se logra del entorno físico a través de los sensores naturales (ojos, oídos, nariz, etc.) contiene incertezas que no pueden ser caracterizadas usando estadística o incluso razonamiento lógico tradicional. Específicamente, la incerteza léxica, trata la imprecisión propia en la mayoría de las palabras humanas, usadas para evaluar conceptos y derivar conclusiones. Con el uso de abstracción y analogías, unas pocas oraciones describirán contextos complejos que son difíciles de modelar con precisión matemática (20, 23, 32).

El estudio exploratorio de los datos imprecisos, presenta restricciones, puestas en evidencia por Ávila-de Hernández *et al.* (2011). Variables que presentan elevado número de valores perdidos, lo cual implica por un lado, pérdida de eficiencia en el análisis, y por otro lado, frente a la posibilidad de que los valores perdidos sigan un patrón no aleatorio, tanto ignorarlos como estimarlos, mediante alguno de los sistemas de imputación, implica un sesgo, inconveniente en la obtención de patrones mediante el análisis factorial (21). Barda *et al.* (2011) señala que los valores medios y la dispersión alertan de posibles deficiencias de los datos observados, en relación con problemas de la realidad estudiada e incluso la propia naturaleza subjetiva de las mediciones (26, 36).

A partir de estos problemas, es indispensable el empleo de métodos robustos, como los que provee el análisis multivariado, con el diseño de una estrategia que permita minimizar el efecto de valoraciones atípicas (outlier) y faltantes.

El amplio intervalo de respuestas del ser humano a los estímulos sensoriales, genera complejidad en la percepción de

la calidad de los alimentos. El empleo de la lógica borrosa permitirá representar y manipular los datos sensoriales subjetivos, con métodos multivariados y univariados conocidos. En este sentido, la definición de los conjuntos difusos que representan los distintos valores de cada variable, es un primer paso en un proceso de razonamiento aproximado en el que se opera sobre la base de percepciones en lugar de medidas de las mismas.

La teoría de la posibilidad, constituyen modelos, basados en la lógica borrosa, que resultan especialmente útiles para tratar con la imprecisión de manera más "natural" y más "humana". Emplean una terminología particular: "Fuzzy" (difuso o borroso); "Crisp" (nítido); "Fuzzification" (borrosificación: convertir un conjunto nítido en borroso); y "Defuzzification" (desborrosificación: convertir un conjunto borroso en un valor "Crisp").

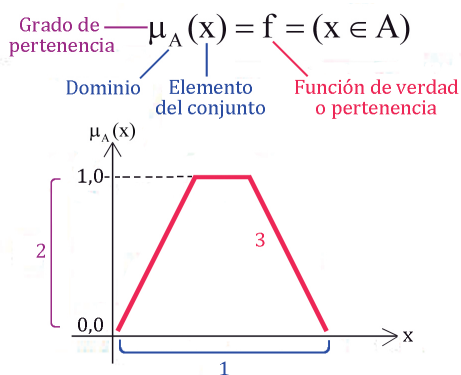
El concepto de la lógica borrosa, surgió en Estados Unidos de América (1965) con Zadeh, y el enfoque difuso explica los valores ambiguos, inexactos, por medio de variables lingüísticas borrosas, donde cada uno de sus valores, se representan con un número borroso. Básicamente, se asignan diferentes valores de verdad a los enunciados y no solamente los clásicos «falso» o «verdadero». Asimismo, ofrece un marco de trabajo natural donde las cantidades borrosas son interpretadas como distribuciones de posibilidad, describiendo los valores del conocimiento incierto o impreciso de conjuntos borrosos (31).

El supuesto de dicha teoría es que existen conjuntos en los que no está claramente determinado si un elemento pertenece o no al conjunto, a veces, un elemento pertenece al conjunto con cierto grado. En el caso de un conjunto de muestras de alimentos dulces, constituye un conjunto difuso, pues no está claro

el límite de dulzura que establece cuál alimento sería dulce o no lo sería, ese límite es borroso y, por tanto, el conjunto que delimita también lo será.

El enfoque difuso permite representar los datos de las escalas sensoriales por medio de variables lingüísticas borrosas, donde cada valor es un número borroso (conjunto difuso). Por ejemplo, la variable lingüística "intensidad aroma" puede tomar los valores lingüísticos "ausente", "muy débil", "débil", "medio, fuerte" y "muy fuerte", tienen un significado semántico y se pueden expresar numéricamente por un grado de verdad entre 0 y 1 (30, 31, 32).

La figura 1, muestra los tres componentes básicos de un conjunto difuso: (1) Eje x (números reales crecientes) que constituyen el Dominio del conjunto difuso; (2) Eje y con valores entre 0 y 1 que significa grado de pertenencia al conjunto; y (3) función de pertenencia (superficie) del conjunto, que conecta un elemento del dominio con su grado de pertenencia. El grado de pertenencia de un conjunto difuso es subjetivo, implica una "definición" más que "medición".



**Figura 1.** Representación de los tres Componentes de un Conjunto difuso.

**Figure 1.** Representation of the three components of a fuzzy set.

Los humanos poseen habilidad para asignar grados de pertenencia a objetos determinados, sin un entendimiento consciente de cómo llegar a ese valor. Otra característica, que aporta a su subjetividad, es que dependerá del contexto. Formalmente, se define a una función de pertenencia como una curva donde un valor de entrada es mapeado a un valor de pertenencia o membresía al conjunto difuso (36).

Transformar los datos sensoriales mediante la matemática borrosa, es un enfoque apropiado para obtener conocimiento preciso a partir de valoraciones subjetivas. Sin embargo, aunque el razonamiento borroso puede tener potencial como método eficaz en el análisis de datos, en la actualidad, hay pocas referencias sobre cómo llevar a cabo la aplicación de la lógica difusa en la sensometría, y no se dispone de criterios y métodos armonizados y normalizados para una evaluación objetiva y reproducible, utilizando la matemática borrosa (8, 10, 20).

En cuanto al análisis de datos borrosos, el Análisis Factorial Multivariado de Datos (AFMD) en la versión de la escuela francesa, surge en la década de los 70, planteando fines menos deterministas que los de la Estadística tradicional. Su objetivo general es la búsqueda de una estructura presente en los datos, en un contexto de tipo más abductivo que deductivo, que revaloriza el rol del individuo estadístico (3).

En la actualidad, las técnicas de AFMD tales como el Análisis Factorial de Correspondencias Simples (ACS) y las pruebas de asociación y validación de tablas de contingencia, presentan salidas gráficas fáciles de interpretar y revelan relaciones no lineales entre las modalidades de diferentes variables que otros métodos no permiten observar. En el análisis de números borrosos es apropiado

el uso de estas técnicas al representar los datos en tablas de contingencia.

El tema que se desarrolla en este trabajo busca exponer un esquema metodológico basado en el enfoque difuso, para el análisis de datos. Con la finalidad de demostrar que el empleo del análisis multivariado a partir de datos difusos que representan la información de los individuos de manera más realista, enriquece el proceso de análisis, ya que constituye un instrumento de observación apropiado que facilita la explotación de datos cuantitativos, cualitativos y también borrosos.

Para lograr este propósito, en el marco de la tesis "Estudio y Aplicación de la lógica borrosa para medir, analizar y evaluar calidad sensorial en alimentos", para optar al título de Doctora en Alimentación, de la carrera en Red Norte de Argentina, se planteó la representación de la incertidumbre a través de la lógica difusa y presentó, a partir del estudio de un caso, la aplicación de técnicas de análisis factorial de correspondencia y técnicas de validación y descripción, en el tratamiento de datos borrosos en estudios de percepción sensorial de tomate triturado, con datos otorgados por la cátedra de Bromatología de la Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo.

## **MATERIALES Y MÉTODOS**

Para establecer un marco de trabajo que resulte flexible y eficiente en el manejo de información imprecisa asociada con el análisis de datos, se representó la incertidumbre a través de la lógica difusa y se aplicó técnicas de análisis multivariante en el tratamiento de datos borrosos, en el ámbito del análisis sensométrico de alimentos de origen vegetal.

Se utilizaron los resultados de pruebas sensoriales realizadas en el 2015, con el

objeto de evaluar la percepción sensorial de tomate triturado, a través de descriptores de aspecto general y consistencia, entre otros. Los datos obtenidos dentro del marco del proyecto SeCTyP "Evaluación sensorial de productos frutihortícolas", fueron provistos por la Cátedra de Bromatología de la Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo, Mendoza. Y se obtuvieron al aplicar una prueba descriptiva, con escalas estructuradas de 5 puntos, a través de un panel de 8 jueces, en siete muestras ("M954", "M632", "M125", "M471", "M580", "M369" y "M293") de tomate triturado con agregado de *Aloe vera* elaborado dentro del proyecto SeCTyP "Uso de gel de *Aloe sp* en alimentos".

La propuesta Metodológica de Análisis de Datos Borrosos, involucra 4 grandes procesos: -1- La codificación y estandarización de los datos medidos; -2- La definición de las variables lingüísticas difusas y su empleo en la transformación de los datos estandarizados en números borrosos representados en una tabla de contingencia; -3- La obtención de un valor borroso colectivo de todo el panel y un valor cuantitativo (crisp) final representativo de la calificación de las muestras evaluadas; -4- aplicación de métodos multivariados para análisis, visualización y obtención de conocimiento sobre los datos borrosos (métodos factoriales de correspondencias para tablas de contingencia y factorial múltiple para estudio de múltiples tablas de contingencia).

### **La codificación y estandarización de datos sensoriales**

Las valoraciones obtenidas de las pruebas sensoriales se presentan en diversas tablas de datos; naturaleza cuantitativa (datos no estructurados tratados como números) o de naturaleza cualitativa (datos estructurados en categorías o etiquetas verbales).

En el caso de aplicación, la tabulación se realizó en una matriz de datos cuantitativos discretos. Se dispuso cada muestra valorada en filas vs. cada descriptor evaluado, en columnas, las valoraciones se representaron por un número, de 1 a 5 para los dos descriptores sensoriales (tabla 1).

Opcionalmente, las puntuaciones en los atributos se normalizan a efectos de eliminar problemas de cálculo originados en la utilización de diferentes escalas. De esta manera se utiliza una misma variable lingüística difusa. Para la estandarización de los datos, se propone utilizar el método de normalización lineal a valores de 0 a 1, usando la ecuación de la recta.

### Transformar las valoraciones a datos borrosos

Los atributos medidos se representan a través de una o varias variables lingüísticas borrosas, lo que requiere seleccionar los descriptores lingüísticos adecuados y su semántica.

**Tabla 1.** Representación esquemática de la Tabla de datos donde se tabularon los datos sensoriales observados de la prueba con muestras de tomate triturado.

**Table 1.** Schematic representation of the data table where the observed sensory test data are tabulated with samples of crushed tomatoes.

Juez	Muestra	Aspecto general	Consistencia
J01	M954	Apreciación (valores de 1 a 5)	
...			
J08			
J01	...		
...			
J08			
J01	M293		
...			
J08			

### Descriptores lingüísticos

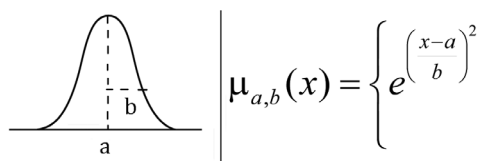
Se elige un conjunto de etiquetas o de términos lingüísticos (conjuntos difusos). Se consideraran  $i$  etiquetas lingüísticas finitas y totalmente ordenadas, como  $C = \{c_0, c_1, \dots, c_i\}$ ,  $c_0 \leq c_1 \leq \dots \leq c_i$ , donde cada etiqueta  $c_i$  representa la valoración lingüística para cada descriptor (se denomina granularidad de la variable). El número de conjuntos difusos, se define por el analista en función de los objetivos de la prueba. Valores típicos de "granularidad" son aquellos impares como 5, 7 ó 9, donde el término medio representa un valor indiferente o punto medio, y el resto se distribuye simétricamente a su alrededor. Sin embargo, las etiquetas y su significado deben ser acordes con las escalas utilizadas en las pruebas, por lo que no siempre el valor central determina la indiferencia, y no siempre las escalas son simétricas (4, 11).

### Semántica de los conjuntos difusos

Este procedimiento define para cada conjunto difuso (términos lingüísticos) una función matemática que representa el grado de (posibilidad o verdad de una percepción) pertenecer a un conjunto borroso.

Los tipos de funciones de pertenencia comúnmente utilizados son: la función Triangular, Trapezoidal y Gaussiana, entre otras (36). Las estimaciones lingüísticas con una escala de medición de la cual se obtiene observaciones numéricas, tienden a seguir una distribución normal y es apropiado el uso de números borrosos Gaussianos.

En un conjunto difuso con función Gaussiana (figura 2, pág. 239), se definen dos parámetros (a y b) que determinan la media o desvío de una campana de Gauss, donde a representa el valor medio central, con total pertenencia al conjunto ( $\mu=1$ ), y b es el desvío que representa la amplitud de la campana.



**Figura 2.** Representación gráfica y parámetros de la Función de Pertenencia Gaussiana. A la derecha se representa el cálculo de  $\mu_{a,b}(x)$  eje vertical

**Figure 2.** Graphical representation and parameters of the Gaussian membership function. To the right is represented calculate  $\mu_{a,b}(x)$  axe vertica.

Otra forma alternativa es la de la Ecuación 1:

$$\mu_{a,b}(x) = \exp[-b \cdot (x-a)^2] \quad (1)$$

En el caso de estudio, para la representación de la intensidad (Aspecto y Consistencia), se definió una misma variable lingüística y correspondiente función de tipo Gaussiana para ambos atributos, con 5 conjuntos difusos, para transformar valores de 1 a 5 respecto de "intensidad o apreciación" (figura 3, pág. 240).

La representación gráfica se efectuó a través de los gráficos XY dispersión (Microsoft Excell) y la herramienta para desarrollo de sistemas de inferencia borrosa de FisPro (Fuzzy Inference System Professional). Dicho sistema es un Software de código abierto que permite crear modelos simples de inferencia difusa y utilizarlos con el propósito de razonamiento, sobre todo para la simulación de un sistema físico o biológico (17).

#### *Borrosificación*

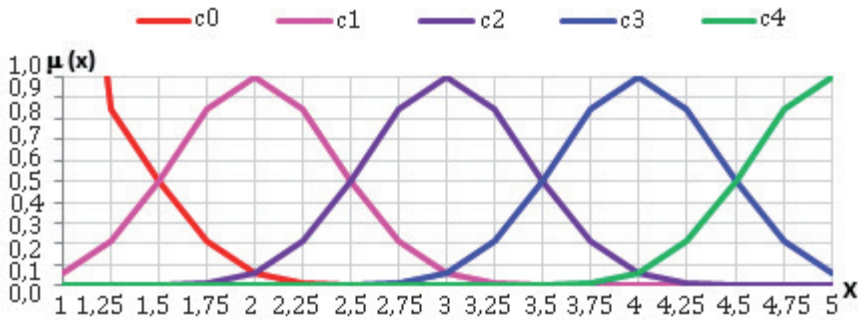
Los datos sensoriales tabulados en punto -1-, se convierte en un número borroso derivado empíricamente. A través

de las ecuaciones de las funciones de pertenencia definidas para las variables lingüísticas en punto -2-, se compara el valor sensorial con la función matemática correspondiente, obteniendo un valor entre cero y uno para cada conjunto difuso. Cada uno de estos valores de posibilidad de asociarse a cada etiqueta lingüística, constituye un número borroso. Los datos borrosificados se representan en una tabla de contingencia, en columna se representa los conjuntos borrosos de cada descriptor, en fila se identifica la percepción para un panelista de cada muestra, la celda representa el valor de posibilidad obtenido. En el caso de estudio, se determinó la expresión de la intensidad de las percepciones de los jueces sobre cada una de las características de las muestras de tomate evaluadas, mediante etiquetas lingüísticas. Para cada valoración de los jueces, se aplicó la Ecuación 1, con los parámetros definidos en la figura 3 (pág. 240, para determinar el valor de pertenencia, de cada conjunto borroso en cada atributo (tabla 2, pág. 241).

#### **Agregación**

En esta etapa, se determina un valor colectivo del panel, para cada característica, aplicando un "operador de agregación" sobre las evaluaciones borrosificadas, y posteriormente un "método de desborrosificación" para calcular un valor preciso (crisp). El procedimiento consiste, en calcular para cada conjunto borroso  $c_i$  la unión  $t(c_i)$  del conjunto difuso.

Se propone un operador de la familia de t-conormas, denominado suma algebraica de los valores de pertenencia  $\mu(c_i)$  de los  $p$  panelistas (ecuación 2, pág. 240). Es considerado apropiado para la agregación de datos sensoriales, siendo un operador probabilístico; este calcula la suma menos el producto de los valores de posibilidad de cada conjunto.



c0: Muy Malo / Inconsistente . c1: Malo / Poco Consistente . c2: Intermedio / Media . c3: Bueno / Consistente  
 c4: Muy Bueno / Muy Consistente  
 c0: Very Bad / Inconsistent . c1: Bad / Little Consistent . c2: Intermediate . c3: Good / Consistent .  
 c4: Very Good / Very Consistent

**Figura 3.** Definición de Variable lingüística difusa de Intensidad (Aspecto General y Consistencia) para borrosificar datos de tomate triturado.

**Figura 3.** Variable fuzzy linguistic definition Intensity (General and Consistency Aspect) for data borrosificar crushed tomatoes.

Se obtiene un valor agregado colectivo del grupo de degustadores, para cada conjunto difuso correspondiente un atributo sensorial.

$$t(c_i) = \sum_1^p \mu(c_i) - \prod_1^p \mu(c_i) \quad (2)$$

donde:

$p$  = número de panelistas (jueces y/o consumidores).

$c_i$  = identifica cada conjunto difuso de la variable lingüística (toma valores desde 0).

Luego, a partir de la agregación  $t(c_i)$ , se identifica la etiqueta lingüística o conjunto difuso colectivo de los  $p$  panelistas; para una granularidad de 5, son  $k = 5$  conjuntos difusos, representados por  $c_i$  etiquetas ( $i$  varía de 0 a 4), se calcula la etiqueta lingüística agregada según la Ecuación 3,

donde  $S$  podrá tomar valores de 0 a 4 (se redondea para obtener un valor discreto), la etiqueta lingüística es la representada por  $l_s$ .

$$S = entero \left( \frac{\sum_1^k t_{c_i} \cdot i}{\sum_1^k t_{c_i}} \right) \quad (3)$$

$k$  = granularidad de la variable lingüística (número de conjuntos difusos).

$i$  = identifica cada etiqueta o conjunto difuso (toma valores de 0 hasta  $k-1$ ).

$t(c_i)$  = agregación del conjunto difuso  $c_i$  (unión).

Por medio de los métodos de desborrosificación, a partir del conjunto de valores borrosos agregados, se obtiene información precisa expresada mediante un valor crisp. Cada método utiliza una fórmula matemática para calcular la salida final.



**Tabla 2.** Representación de los cálculos de borrosificación, agregación, valor borroso y desborrosificación de los valores de aspecto y consistencia de la muestra M632 de tomate triturado.

**Table 2.** Representation fuzzification calculations, aggregation, and defuzzification fuzzy value of the securities in appearance and consistency of the M632 sample of crushed tomatoes.

Juez	Muestra	Aspecto General	AG-c0	AG-c1	AG-c2	AG-c3	AG-c4	Consistencia	CO-c0	CO-c1	CO-c2	CO-c3	CO-c4
J01	M632	5	0,000	0,000	0,000	0,064	1,000	4	0,000	0,000	0,064	1,000	0,064
J02	M632	1	1,000	0,064	0,000	0,000	0,000	5	0,000	0,000	0,000	0,064	1,000
J03	M632	3	0,000	0,064	1,000	0,064	0,000	3	0,000	0,064	1,000	0,064	0,000
J04	M632	4	0,000	0,000	0,064	1,000	0,064	2	0,064	1,000	0,064	0,000	0,000
J05	M632	5	0,000	0,000	0,000	0,064	1,000	4	0,000	0,000	0,064	1,000	0,064
J06	M632	5	0,000	0,000	0,000	0,064	1,000	5	0,000	0,000	0,000	0,064	1,000
J07	M632	5	0,000	0,000	0,000	0,064	1,000	4	0,000	0,000	0,064	1,000	0,064
J08	M632	4	0,000	0,000	0,064	1,000	0,064	1	1,000	0,064	0,000	0,000	0,000
	<b>M632</b>	<b>3,97</b>					<b>4,13</b>	<b>c3</b>					<b>c2</b>

$$\mu_{a,b}(x) = \exp^{-b \cdot (x-a)^2}$$

$$S = \text{entero} \left( \frac{\sum_{i=1}^k t_{ci} \cdot i}{\sum_{i=1}^k t_{ci}} \right)$$

0,000 = Borrosificación (Ecuación 1)

**c3 / c2** = Valor Borroso (Ecuación 3)

$$t(c_i) = \sum_{i=1}^p \mu(c_i) - \prod_{i=1}^p \mu(c_i)$$

$$V_{\text{CRISP}} = \frac{\sum_{i=1}^k \left( \frac{t(c_i)}{b^2} \cdot a_i \right)}{\sum_{i=1}^k \left( \frac{t(c_i)}{b^2} \right)}$$

**4,13** = Agregación (Ecuación 2)

**3,97 / 3,49** = DesBorrosificación (Ecuación 4)

La bibliografía sugiere diversas alternativas de operadores de desborrosificación (36). En el caso de usar funciones de tipo gaussianas, se propone implementar el método de la Altura modificado; se divide la agregación  $t$  del conjunto difuso  $c_i$ , por el desvío (parámetro  $b$ ) elevado al cuadrado, finalmente para obtener el valor preciso correspondiente se utiliza la Ecuación 4.

$$V_{\text{CRISP}} = \frac{\sum_1^k \left( \frac{t(c_i)}{b^2} \cdot a_i \right)}{\sum_1^k \left( \frac{t(c_i)}{b^2} \right)} \quad (4)$$

donde:

$k$  = granularidad de la variable lingüística (número de conjunto difuso)

$i$  = identifica a un conjunto difuso (etiqueta lingüística de 0 a  $k-1$ )

$a_i$  = centro de gravedad de cada conjunto  $c_i$ , y  $b^2$  = desvío al cuadrado

$t(c_i)$  = unión de los valor de pertenencia de los  $p$  panelista para un conjunto  $c_i$ .

Para el caso de aplicación, a través de las valoraciones borrosificadas, se calculó un valor colectivo, para cada característica, a través de la aplicación del operador de agregación "suma algebraica" (Ecuación 2, pág. 240). Los resultados por muestra, se traspusieron en una tabla de contingencia.

A partir de la Ecuación 3 (pág. 240), para cada etiqueta  $l_s$ , se asignó un número borroso según los parámetros ( $a$  y  $b$ ) de la variable lingüística.

A partir de las valoraciones borrosificadas y agregadas, se construyeron las tablas de contingencia apropiadas, con fines de continuar la estrategia de análisis propuesta en el siguiente procedimiento para la explotación de los datos difusos obtenidos (tabla 2, pág. 241).

### Estrategia metodológica para el análisis de datos borrosos

Se establecen dos sistemas de tablas de contingencia, a partir de los números borrosos agregados en la unión y las valoraciones borrosificadas.

Todos los datos son estandarizados mediante el método de normalización lineal en valores de 0 a 100, de esta manera se representa porcentaje de posibilidad que asocia un elemento en fila con un elemento columna. La visualización de asociaciones se realiza a través de un Análisis factorial de correspondencias simples (AFCS) (6, 14, 15, 18).

La interpretación, descripción y validación, se efectúa por las pruebas del  $\chi^2$ , y prueba exactas de Fischer y el valor de prueba *Valor de Test* (5, 6, 9, 19, 24, 29). Para la puesta en práctica de la estrategia de análisis factorial y pruebas de asociación, se utilizó el software de libre distribución para uso académico DTM. Minería de Datos y Textos, que trata la estadística exploratoria multidimensional de datos complejos, incluyendo datos numéricos y textuales (18); y el módulo de análisis sensorial del software comercial Xlstat, bajo la licencia para investigar 2015 (30).

### RESULTADOS Y DISCUSIÓN

A partir del cálculo del valor colectivo borroso, para cada característica, se construyeron las tablas de contingencia para cada atributo, de manera de comparar las muestras de tomate triturado.

En la tabla 3 (pág. 243), los datos asocian las muestras (filas) con los conjuntos difusos de las variables lingüísticas utilizadas en la borrosificación (columnas), y representa el porcentaje de posibilidad, una muestra estará caracterizada por un conjunto difuso en particular por la mayor posibilidad de pertenecer al mismo.

**Tabla 3.** Tablas de contingencia representativas de la agregación difusa de todas las muestras para cada descriptor. Datos agregados y normalizados que indican porcentaje de posibilidad de pertenecer a un conjunto difuso. Resaltado en celeste, asociación de fila-columna con alto nivel de significación estadística (alfa 0,05) Dos columnas con el valor crisp agregado de cada atributo.

**Table 3.** Contingency tables representative of fuzzy aggregation of all samples for each descriptor. aggregated and normalized data indicating percent chance of belonging to a fuzzy set. Highlighted in blue, row-column association with high statistical significance (alpha 0.05) Two columns with crisp added value of each attribute.

Agregación	AG-c0	AG-c1	AG-c2	AG-c3	AG-c4	CO-c0	CO-c1	CO-c2	CO-c3	CO-c4	Aspecto general	Consistencia
M954	0	0	0	7	100	0	0	0	7	100	4,93	4,93
M632	14	2	16	33	59	15	16	18	46	31	3,97	3,49
M125	2	29	17	19	58	46	47	18	15	1	3,81	2,04
M471	43	5	30	31	16	0	1	19	74	33	2,79	4,10
M580	3	44	17	5	57	16	30	19	45	17	3,56	3,13
M369	0	4	57	7	57	16	29	5	58	18	3,94	3,25
M293	3	47	60	5	14	100	21	1	0	0	2,85	1,19

Los resultados de la prueba exacta de Metha *et al.* (1983) y Connelly (2016), para caracterizar cada atributo, permiten resaltar en la tabla, en celeste, el nivel de significación estadística de la asociación, para cada descriptor respecto a la muestra-conjunto posiblemente asociado.

Los resultados de las pruebas estadísticas se expresan indicando la probabilidad (p.value) de que un resultado específico pueda ocurrir por casualidad y no sea una diferencia real. Si un resultado ocurre por casualidad 5 de cada 100 veces, se dice que la probabilidad es de 0,05.

Por lo general, un resultado estadístico se considera significativo, solamente si tiene una probabilidad de 0,05 o menos. No es posible leer la tabla directamente, en este caso el análisis simultáneo de todas las muestras a través del Análisis factorial de Correspondencias Simple (AFCS), permite visualizar en un mismo mapa perceptual las mismas de manera comparativa (figura 4, pág. 244).

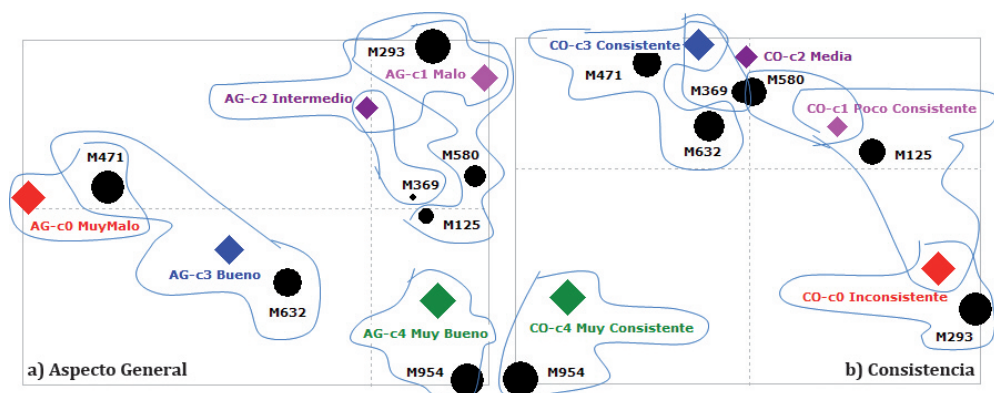
La proximidad de puntos, muestra la asociación por mayor porcentaje de posibilidad. Si una muestra se encuentra

entre dos características (muestra-conjunto difuso), esto indica que el mayor porcentaje de posibilidad se da en forma más o menos pareja entre esos dos conjuntos, por lo tanto ambas son características de esa muestra.

Visualmente el tamaño indica la calidad de representación en el mapa. La interpretación de un punto pequeño cercano al centro, indica que el mismo no está asociado a una característica en particular debido a presentar una distribución de posibilidad muy pareja para todas las características sin poder asociarla a alguna en especial.

Con alto nivel de significación, es posible validar la tendencia de que la muestra M471 es la de peor aspecto (c0), no así para la consistencia, la muestra más inconsistente es la M293.

Sin embargo, es posible observar que la muestra M471 también se acerca a un aspecto bueno, esto se debe a que la posibilidad es mayoritaria casi para ambos conjuntos (características), lo que refleja problema de entrenamiento de los jueces.



**Figura 4.** Mapa perceptual de la agregación borrosa de todas las muestras, comparativo del a) aspecto general y b) consistencia. Proximidad representa asociación, tamaño nivel de significación.

**Figure 4.** Perceptual map of fuzzy aggregation of all samples, comparison of a) general appearance and b) consistency. Proximity represents association, size and quality of representation and significance level.

Visualmente se puede observar en el plano de la figura 4a, cómo la muestra M471 se encuentra en el medio del aspecto muy malo y bueno, siendo su distribución de probabilidad mayoritaria para ambas características.

En cuanto a Consistencia no es posible validar una tendencia para la muestra M632, debido a la gran dispersión entre los jueces para valorar esta muestra, sin embargo se puede observar visualmente en el plano de la figura 4b, su proximidad al conjunto C<sup>3</sup>, lo que indica una muestra Consistente.

Para la interpretación de los mapas resulta de gran importancia el uso del "Valor Test", que se calcula a partir de la prueba exacta de Fisher, siendo la tabla normalizada a valores de 0 a 100, los resultados coinciden para la prueba del Chi<sup>2</sup>; (este valor refleja para valores mayores a 1,95 un nivel de significación de  $\alpha = 0,05$ , y para mayores de 2,58 un nivel de  $\alpha = 0,01$ ).

En la tabla 4a (pág. 245), se sintetiza el perfil difuso de cada muestra, el valor de test indica la validez en la asociación de un

conjunto difuso a la muestra. En el caso de la muestra M471 se observa la asociación de un aspecto general tanto muy malo (c0) como bueno (c3), lo que refleja la división del panel y la imposibilidad de discriminar el aspecto de dicha muestra por la falta de una referencia. Sin embargo, el valor de test indica la tendencia de que el grupo mayoritario de jueces consideran la muestra de muy mal aspecto (AG-c0).

La tabla 4b (pág. 245), refleja las medias aritméticas de las muestras y el valor de test obtenidos a partir de una prueba de varianza de un factor a través de Fisher, las medias características. A pesar del nivel de significación estadístico, es posible observar que las medias están distorsionadas y no muestran la realidad, afectadas por la alta dispersión del panel de jueces en cuanto al aspecto de las muestras. Siguiendo con el ejemplo, la muestra M471 se caracteriza con alto nivel de significación por una media esperada de 2,8 en aspecto, muy alejado de la realidad que se expone en la tabla 4a (pág. 245).

**Tabla 4.** Comparativa del perfil difuso y clásico de las muestras de tomate. a) Conjuntos difusos asociados por muestras a partir del porcentaje de posibilidad validado por los valores de test; b) media aritmética esperada de los atributos sensoriales medidos por muestra y validada por el valor de test ( $V_{Test} > 2 \alpha = 0,05$ ).

**Table 4.** Comparison of fuzzy and classical profile samples tomato. a) Fuzzy sets associated with samples from the percentage of possible values validated by test; b) expected arithmetic average of the sensory attributes measured per sample and validated by the test value ( $\sqrt{V_{Test}} > 2 \alpha = 0.05$ ).

a)				b)			
Conjunto difuso	% posibilidad	Valor-Test		Variables características	Media esperada	Valor-Test	
<b>M954</b>				<b>M954</b>			
CO-c4	47	14,26	***	Aspecto general	4,9	29,91	***
AG-c4	47	9,05	***	Consistencia	4,9	24,96	***
<b>M471</b>				<b>M471</b>			
AG-c0	17	9,44	***	Consistencia	4,1	14,75	***
CO-c3	29	6,77	***	Aspecto general	2,8	-22,86	***
AG-c3	12	3,85	***	<b>M632</b>			
<b>M632</b>				Aspecto general	4,0	7,85	***
AG-c3	13	4,37	***	Consistencia	3,5	5,49	***
<b>M369</b>				<b>M369</b>			
AG-c2	23	5,47	***	Aspecto general	3,9	7,13	***
CO-c3	23	4,01	***	Consistencia	3,3	1,94	**
<b>M580</b>				<b>M580</b>			
AG-c1	17	5,51	***	Aspecto general	3,6	-2,89	***
CO-c1	12	2,01	**	<b>M125</b>			
<b>M125</b>				Aspecto general	3,8	3,78	***
CO-c1	19	5,48	***	Consistencia	2,0	-16,39	***
CO-c0	18	3,37	***	<b>M293</b>			
AG-c1	12	2,41	**	Aspecto general	2,9	-21,01	***
<b>M293</b>				Consistencia	1,2	-29,26	***
CO-c0	40	13,30	***				
AG-c1	19	6,22	***				
AG-c2	24	5,92	***				

## CONCLUSIONES

Los resultados obtenidos, constituyen un instrumento de mejora a la evaluación sensorial, optimizando la obtención de perfiles de calidad sensorial del alimento.

La descripción de propiedades sensoriales de los productos es complicada, incluso para los expertos, ya que

pueden utilizar diferentes variables y ponderaciones. Sin dudas, inclusive, describen herramientas complementarias que aportan mucha información en cuanto al estudio del producto vegetal en fresco, mínimamente procesado o procesado. Entre ellas, un análisis de la dinámica y

evolución de atributos relacionados con la calidad sensorial, estudio de factores influyentes y modelización de su comportamiento a través de la lógica borrosa. Lo propuesto se considera una base inicial de futuros trabajos generadores de nuevas metodologías instrumentales para la determinación de atributos de calidad, adaptación de tecnologías existentes, y desarrollo de nuevos dispositivos para uso en laboratorio o en línea.

Quedan en evidencia las propiedades ventajosas del uso de números borrosos y métodos multivariados: el proceso de borrosificación y agregación borrosa, es apropiado para calcular similitudes cuando las características que se miden involucran variables de diferente naturaleza (cuantitativa, cualitativa, binaria, textual). Los "números borrosos" son capaces de captar por sí mismos la vaguedad lingüística de palabras y frases comúnmente aceptadas en los estudios sensométricos, como "bastante intenso", "algo aceptable" o "ligero cambio". Se logró representar información que casi siempre es poco precisa, de manera más apropiada y que facilita la convergencia del lenguaje hacia significados compartidos. Permite lograr mayor precisión y detalle en el análisis multivariado de datos, sin que las valoraciones atípicas, afecten los resultados. También, la metodología

propuesta ayuda a determinar en qué atributos hay mayores discrepancia, lo cual permite planificar el entrenamiento del panel concentrándose en los descriptores relevantes. Otra característica es que es posible trabajar con datos faltan observaciones de algunas variables, sin implicar que se suprima de todo el vector que representa a la unidad muestral. Finalmente, los resultados obtenidos confirma que los "valores test" basados en la prueba de Fisher, constituyen una herramienta rápida para la interpretación y validación de los resultados de análisis factoriales.

Se puede afirmar que el empleo del Análisis Multivariado a partir de valoraciones borrosas, plantea un nuevo paradigma en la sensometría, y enriquece el proceso de análisis sensorial de un alimento constituyendo un instrumento de observación apropiado, facilitando la explotación de datos sensoriales. Queda abierta la discusión sobre la metodología planteada y se requiere profundizar acerca de las dispersiones de las funciones de pertenencia, las correlaciones de dichas funciones, etc. Lo dicho lleva a proponer nuevas líneas de investigación como la orientada a la forma de estimar de manera objetiva, los parámetros de la variable lingüística en función de las valoraciones, diseñando variables lingüísticas a la medida de cada estudio.

## BIBLIOGRAFÍA

1. Ávila de Hernández, R. M.; González-Torrevilla, C. C. 2011. La evaluación sensorial de bebidas a base de fruta: una aproximación difusa. *Universidad, Ciencia y Tecnología*. 15(60): 171-182.
2. Barda, N. 2011. Análisis sensorial de los alimentos. *Fruticultura & Diversificación*. Disponible en: <http://www.biblioteca.org.ar/libros/210470.pdf>
3. Bécue-Bertaut, M., Salazar, B. C.; Vinué, P. F. 2009. Extensión del análisis factorial múltiple jerárquico a datos textuales: Aplicación en sensometría. In XXXI Congreso Nacional de Estadística e Investigación Operativa; V Jornadas de Estadística Pública: Murcia, 10-13 de febrero de 2009: Libro de Actas. p. 27.
4. Bonissone, P. P.; Decker, K. S. 2013. Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity. arXiv preprint arXiv: 1304.3425.

5. Césari, M.; Rodríguez, D.; Rancan, C.; Merlino, H.; Britos, P. V.; García Martínez, R. 2008. Métodos iconográficos de observación, exploración y comunicación aplicados a la minería de textos. In X Workshop de Investigadores en Ciencias de la Computación. 5 p.
6. Césari, R.; Césari, M. 2012. Material pedagógico de cursos de postgrado en métodos para analizar datos como soporte a la investigación en ingeniería, Capítulo 03: "Las técnicas del análisis multivariado de datos" y Capítulo 04 "Procedimiento de descripción y validación estadística", Maestría en Medio Ambiente de la UNCuyo y Especialización en Redes de Datos de la FRM, UTN, Mendoza, Argentina.
7. Connelly, L. M. (2016). Fisher's exact test. *MedSurg Nursing*. 25(1): 58-60.
8. Debjani, C.; Das, S.; Das, H. 2013. Aggregation of sensory data using fuzzy logic for sensory quality evaluation of food. *Journal of food science and technology*. 50(6): 1088-1096.
9. Díaz, S. P.; Fernández, S. P. 2004. Asociación de variables cualitativas: el test exacto de Fisher y el test de McNemar. *Cuadernos de atención primaria*. 11(5): 304-308.
10. Espinilla, M.; Martínez, L.; Pérez, L. G. 2008a. A linguistic multigranular sensory evaluation model for olive oil. *International Journal of Computational Intelligence Systems*. 1(2): 148-158.
11. Espinilla, M.; Martínez, L.; Pérez, L. G.; Liu, J. 2008b. Modelo de evaluación sensorial con información lingüística multigranular para el aceite de oliva. XIV Congreso Español sobre Tecnologías y Lógica fuzzy. Cuencas Mineras (MieresLangreo). p. 249-255.
12. Ferreira, M. S.; Santos, M. C.; Moro, T. M.; Basto, G. J.; Andrade, R. M.; Gonçalves, É. C. 2013. Formulation and characterization of functional foods based on fruit and vegetable residue flour. *Journal of food science and technology*. 52(2): 822-830.
13. Fiszman, S. 2005. Análisis sensorial aplicado a la evaluación de las frutas y hortalizas cortadas. nuevas tecnologías de conservación y envasado de frutas y hortalizas. Centro de Investigación en Alimentación y Desarrollo, AC, Hermosillo. México.
14. Greenacre, M. 2007. Correspondence analysis in practice. Second Edition CRC Press, 7 may. 2007. 296 p.
15. Greenacre, M. 1994. Correspondence analysis and its interpretation. *Correspondence analysis in the social sciences*. 3-22.
16. Guerrero, L.; Guardia, M. D. 1998. Evaluación de la fiabilidad de un panel de cata. III Jornadas de Análisis sensorial. Valdediós, Villaviciosa, Asturias.
17. Guillaume, S.; Charnomordic, B. 2012. Fuzzy inference systems: An integrated modeling environment for collaboration between expert knowledge and data using FisPro. *Expert Systems with Applications*. 39(10): 8744-8755.
18. Langrand, C.; Pinzón, L. M. 2009. Análisis de datos: métodos y ejemplos. C. S. Perdomo (Ed.). Editorial Escuela Colombiana de Ingeniería Julio Garabito, Bogotá, Colombia. 388 p.
19. Lebart, L. 2007. DTM. Data and Text Mining, Software Estadística Exploratoria Multidimensional para datos complejos que incluyen datos numéricos y textuales. Disponible en: <http://ses.telecom-paristech.fr/lebart/>
20. Liébana, F. J. E. 2013. Modelo de evaluación sensorial para el aceite de oliva basado en una nueva escala lingüística no balanceada. Trabajo Tutelado de Iniciación a la Investigación. Universidad De Jaen. Disponible en: [http://sinbad2.ujaen.es/cod/archivosPublicos/dea/TTII\\_Estrella.pdf](http://sinbad2.ujaen.es/cod/archivosPublicos/dea/TTII_Estrella.pdf)
21. Little, R. J.; Rubin, D. B. 2014. Statistical analysis with missing data. John Wiley & Sons.
22. Loyola, N.; Acuña, C.; Fuentes, D.; Arriola, M. Production of cured, voided meat sausage with apple pomade added (*Malus domestica*) and its impact in nutritional and sensorial values. *Revista de la Facultad de Ciencias Agrarias*. Universidad Nacional de Cuyo. Mendoza. Argentina. 50(1): 249-262.
23. Martínez, L. (2007). Sensory evaluation based on linguistic decision analysis. *International Journal of Approximate Reasoning*. 44(2): 148-164.
24. Mehta, C.; Patel, N. R. 1983. A network algorithm for performing Fisher's exact test in  $r \times c$  contingency tables. *J. Amer. Statist. Assoc.* 78: 427-434.
25. Pattee, H. E. 2014. Evaluation of quality of fruits and vegetables. Springer, 14 nov. 2014. 428 p.

26. Ponce-Donoso, M.; Vallejos-Barra, O. 2016. Valoración de árboles urbanos, comparación de fórmulas. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. Mendoza. Argentina.* 48(2): 195-208.
27. Stone, H.; Bleibaum, R.; Thomas, H. A. (Eds.). 2012. *Sensory Evaluation Practices.* Academic Press, 4th Edition. 448 p.
28. Torres, J. G.; Busto, O.; Mestres, M. 2011. Nariz electrónica. *Aplicaciones enológicas. ACE: Revista de enología.* 126: 3.
29. Upton, G. J. 1992. Fisher's exact test. *Journal of the Royal Statistical Society. Series A (Statistics in Society).* 395-402.
30. Xlstat versión 2015, licencia para investigación. add-in de análisis estadístico para Microsoft Excel desarrollada por Addinsoft 1996-2016. Disponible en: [www.xlstat.com/es/products/xlstat-pro/](http://www.xlstat.com/es/products/xlstat-pro/)
31. Zadeh, L. A. 1965. Fuzzy sets. *Information and control.* 8(3): 338-353.
32. Zadeh, L. 2014a. *Computing with Words: Principal Concepts and Ideas.* Springer Publishing Company.
33. Zadeh, L. A. 2015. The information principle. *Information Sciences.* 294: 540-549.
34. Zadeh, L. A.; Fu, K. S.; Tanaka, K. (Eds.). 2014b. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes: Proceedings of the US-Japan Seminar on Fuzzy Sets and Their Applications, Held at the University of California, Berkeley, California, July 1-4, 1974.* Academic press.
35. Zimek, A.; Schubert, E.; Kriegel, H. P. 2012. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal.* 5(5): 363-387.
36. Zimmermann, H. J. 2014. *Fuzzy Set Theory-and Its Applications* 3. Springer Publishing Company, 435 p.