



EXPLORACIÓN DE TÉCNICAS DE INTELIGENCIA COLECTIVA EN ANÁLISIS SENSORIAL

TRABAJO FIN DE MÁSTER

Curso: 2017/18

**Alumno: Indalecio Marcos Aldea
Tutor: Encarnación Fernández Fernández
Cotutor: David Orden Martín**

Máster en Calidad, Desarrollo e Innovación de Alimentos
E.T.S. Ingenierías Agrarias, Campus de la Yutera (Palencia)
Universidad de Valladolid

Índice de contenidos

1. Introducción	4
2. Objetivo	9
3. Materiales y métodos	9
3.1. Muestras.....	9
3.2. Pruebas.....	10
3.3. Jueces	11
3.4. Lugar de realización.....	12
3.5. Análisis de datos.....	12
3.5.1. Análisis de datos desde el punto de vista sensorial	12
3.5.2. Análisis de datos desde el punto de vista de la inteligencia colectiva.....	13
4. Resultados y discusión	13
4.1. Galletas.....	13
4.2. Vinos.....	20
5. Conclusión	26
6. Agradecimientos	27
7. Bibliografía	27

Resumen: El presente estudio tiene por objetivo la aplicación de técnicas de inteligencia colectiva en el campo del análisis sensorial. La inteligencia colectiva (IC) se describe como la inteligencia inherente a un grupo heterogéneo de personas, su aplicación en numerosos campos se viene investigando, con buenos resultados. Sin embargo su aplicación al análisis sensorial es todavía novedosa y necesita una mayor investigación.

Este ensayo se considera una primera aproximación inicial en la que se han evaluado nueve muestras de galletas tipo cookies y nueve muestras de vino tinto roble de diferentes Denominaciones de Origen mediante pruebas de consumo con jueces no entrenados. Los resultados estadísticos se han obtenido a partir de distintas pruebas; ANOVA, Friedman y Kruskal-Wallis; mientras que los resultados de inteligencia colectiva se han obtenido a partir de métodos de agregación y algoritmos.

Los resultados obtenidos son prometedores y establecen precedentes en la aplicación de la I.C. en el campo de análisis sensorial. No obstante es necesaria una mayor investigación en la materia para su uso estandarizado.

Palabras clave: Inteligencia colectiva, Análisis sensorial, Vino tinto, Galleta cookie

Abstract: The present study aims to apply collective intelligence techniques in the field of sensory analysis. Collective intelligence is described as the inherent intelligence in a heterogeneous group of people; its application has been investigated in fields such as software development, direct democracy and social relationships. However, its application to sensory analysis is still novel and needs further investigation.

This essay claims to be a first approach nevertheless has very promising results. Nine samples of D.O wines and cookies had been tested by untrained judges. Every set of samples has been tested with statistics methods; ANOVA, Friedman's Test, Kruskal-Wallis H; and with aggregation methods; simple aggregations and algorithms.

The results obtained are encouraging and establish precedents in the application of the I.C. in the field of sensory analysis. However, further research in the field is necessary for its standardized use.

Keywords: Collective intelligence, Sensory analysis, Red wine, Cookie

1. Introducción

La inteligencia colectiva (IC) es la inteligencia compartida que surge de la colaboración, los esfuerzos colectivos y la competencia de un conjunto de personas a la hora de resolver problemas que les afectan (Schut, 2010).

Se puede entender como una propiedad emergente entre las personas y los métodos de procesado de información, que contribuyen a un intercambio de poder y conocimiento desde el individuo al colectivo (Lévy, 1994). Se aplica a la identificación de problemas y a la adopción de ideas o prácticas que sirvan de soluciones parciales o totales de los mismos.

El primer antecedente de este pensamiento se encuentra en 1785 con Marqués de Condorcet, el cual afirmaba en su Teorema del Jurado que si cada miembro de un grupo tiende a votar decisiones que ciertamente son beneficiosas, la probabilidad de que el mayor nº de personas de un grupo vote la elección correcta se incrementa con el aumento del nº de miembros del grupo (Landemore 2012).

Posteriormente Engelbart (1962), conectó la inteligencia colaborativa con la eficacia de las organizaciones y predijo que un aumento de la inteligencia humana se traduciría en un efecto multiplicador en la capacidad del grupo para la resolución de problemas. Tres personas trabajando juntos serán más efectivas resolviendo problemas complejos que la suma de los tres por separado.

Más recientemente, Lévy (2004) ha definido a la inteligencia colectiva como una inteligencia repartida entre todos los miembros de un grupo, en la que nadie lo sabe todo pero, en la que todos sus miembros tienen algo que aportar. La información que generan dichos grupos se valoriza y coordina constantemente en tiempo real, lo cual permite a los colectivos sociales interactuar en campos dinámicos y compartidos.

Esta inteligencia según Lévy (2004), conduce a una movilización de competencias, reconociendo la diversidad de conocimientos y la verdadera identidad social. Por lo que acepta la existencia de diferencias en la inteligencia de los distintos individuos sin negar la validez de ninguna.

En general, se resume como una propiedad que nace del conjunto de personas y los métodos de procesado de información. Y contribuye a una reciprocidad del poder y conocimiento entre el individuo y el colectivo.

Los sistemas de inteligencia colectiva son complejos por naturaleza, y se pueden entender como el resultado de las interacciones entre:

- El conocimiento y la manipulación de datos.
- El tratamiento y la transmisión de datos.
- Las instituciones y autoridades jerarquizadas, que dirigen al grupo y son capaces de aprender de él.

La toma de decisiones más favorables del grupo depende de las sinergias que se establezcan entre estos tres factores.

Habitualmente se ha relacionado a la inteligencia colectiva con ciencias como la sociobiología, la ciencia política y la informática. Dentro del campo de la informática destaca en el desarrollo de videojuegos, como World of Warcraft o Second Life, páginas web, como Wikipedia, y organizaciones como Creative Commons. Estas actividades se basan en estrategias de Crowdsourcing, en donde a través de una convocatoria abierta en la red todos los usuarios tienen la oportunidad para colaborar en la realización de un trabajo. El resultado final es la suma del trabajo colectivo de un grupo de personas que colaboran para obtener soluciones ante un mismo problema.

Estos sistemas están auto-organizados y se caracterizan por ser altamente eficientes y sumamente inteligentes. Se relacionan con el consenso y la deliberación como motores fundamentales de la toma de decisiones. Donde todos sus participantes aportan y aprenden algo, formando redes en las que, a pesar de que todos los individuos presentan puntos ciegos, éstos son cubiertos por el conocimiento del resto de sus compañeros.

Los sistemas inteligentes y autogobernados a pesar de lo que en un principio se pueda pensar, no son exclusivos de los seres humanos y se encuentran en la naturaleza de forma habitual. Algunos ejemplos de sistemas auto-organizados se pueden encontrar en animales que individualmente no se caracterizan por tener gran inteligencia como pueden ser colonias de insectos, bandadas de pájaros o bancos de peces entre otros.

Sin embargo, al igual que los seres humanos, estos sistemas de animales se caracterizan por ser sumamente complejos y altamente adaptativos ante ambientes desconocidos, se organizan de manera autónoma, pudiendo desarrollar comportamientos “emergentes” (Schut, 2010). Estos comportamientos se caracterizan por ser complejos, dar sensación de inteligentes y aparecen de forma espontánea e imprevista como consecuencia de una serie de acciones simples.

Por lo tanto, aunque la formación de grupos inteligentes no es un rasgo exclusivamente humano, la complejidad de las relaciones que se forman entre los miembros del grupo sí lo es. Cuando se trata de animales, en ocasiones estas relaciones se reducen meramente a reflejos (Lévy, 2015).

El nivel de complejidad de las relaciones humanas ha propiciado un aumento del nivel de inteligencia colectiva, el cual se apoya en los tres factores anteriormente mencionados; conocimiento de datos, transmisión de información y formación de estructuras que canalicen y aprovechen la información.

Esta situación se observa en cualquier sociedad actual, en la cual existe un lenguaje que favorece el aprendizaje y facilita la transmisión de ideas. Y que cuenta con una tecnología que permite dicha transmisión, con independencia de la distancia y el tiempo que separe al emisor del receptor.

Dicha tecnología puede entenderse como los sistemas de escritura tradicionales y los nuevos sistemas informáticos. Lo cual además, ha permitido desarrollar organizaciones capaces de canalizar toda la información generada por los grupos para continuar mejorando en el tiempo, cerrando así un bucle que permite una mejora continua de la IC en el tiempo.

Un aspecto fundamental para que se desarrolle esta mejora continua reside en la capacidad de aprendizaje humana. Este se caracteriza por no sustituir el conocimiento anterior, sino que lo preserva, lo profundiza y lo complementa, creando un sistema de capas cada vez más poderoso (Soberón, 2016).

De esta forma, cuanto mayor sea el desarrollo tecnológico y organizativo de una sociedad, mayor será el índice de inteligencia colectiva de la misma.

A pesar de que existe una relación entre la inteligencia de los individuos y la inteligencia del grupo, no siempre el grupo de personas de mayor coeficiente intelectual va a establecer el grupo de mayor inteligencia colectiva.

Para alcanzar el máximo nivel organizativo dentro de un grupo es necesario estructurarlo de forma correcta y equilibrada, de forma que todos los individuos puedan desarrollar el máximo nivel potencial. Para ello se debe mantener un equilibrio entre la diversidad y la experiencia de los participantes.

La diversidad puede aportar diferentes perspectivas para la resolución de un problema, pero no ayudará en nada si los participantes carecen de experiencia y son completamente ignorantes en los temas que se tratan.

Por estas razones se hace necesario establecer un método apto para la medición de la eficiencia del grupo en la resolución de problemas. La medición de la inteligencia colectiva se puede emplear para predecir el desempeño del grupo ante tareas de diferente dificultad. Este tipo de marcadores ya se aplican para la inteligencia individual general y se conoce como "Factor G".

El marcador que se aplica exclusivamente a la IC es el "Factor C" (Woolley *et al.*, 2010), se define como la capacidad general de un grupo para realizar un amplio rango de tareas así como su habilidad cognitiva como grupo. También permite la comparación de actuaciones de distintos grupos ante una misma tarea, e incluso, permite la predicción a cerca de la actuación del grupo ante tareas similares en un futuro.

Matemáticamente, el "Factor C" y el "Factor G" son variables que resumen las correlaciones positivas entre diferentes tareas, suponiendo que el rendimiento en una tarea es comparable con el rendimiento en otras tareas similares. (Kamphaus, Winsor, Rowe y Kim, 2005)

Este "Factor C" es un indicador de la inteligencia grupal, pero no debe asociarse con el coeficiente intelectual promedio de los miembros del grupo, ni con el coeficiente más elevado de alguno de sus miembros.

La inteligencia individual se ha demostrado que está influenciada tanto por genética como por el medio ambiente. El hecho de que no todos los grupos actúen con la misma eficiencia indica que la inteligencia individual de los miembros y el "Factor C" presentan cierto grado de correlación.

De acuerdo con Woolley y Malone (2011) ni la cohesión, ni la motivación, ni la satisfacción de los grupos están correlacionadas con el "Factor C" como en un

principio se podía pensar, sino que existen otros factores cuya correlación es más significativa:

- La participación distribuida, mide el grado de implicación de los miembros en función del nº de veces que aportan información al grupo.
- La sensibilidad social promedio de los miembros, que se puede traducir en la suma de empatía y sensibilidad.
- La proporción de mujeres en el grupo, se cree una variable del factor anterior ya que por norma general suelen presentar mayor grado de empatía y sensibilidad (Wolley y Malone 2011).

Estudios posteriores de Aggarwal y Wolley (2014), además de identificar al "Factor C" confirmaron que éste podía predecir el rendimiento grupal en diversas tareas. Y no solo eso, sino que además, descubrieron que los integrantes de los grupos de mayor inteligencia colectiva no obtenían los mejores resultados en pruebas individuales inicialmente, pero sí eran los que más mejoraban cuantitativamente, demostrando de esta forma que los equipos de mayor inteligencia colectiva tienen una capacidad de aprendizaje más rápida y mejor.

No existe unanimidad a cerca del número máximo de personas que componen un tamaño de grupo efectivo. Algunos autores como Green (2015), han demostrado que la eficiencia del grupo es mayor cuanto más reducido sea este. De acuerdo a sus conclusiones, el tamaño máximo del grupo debe ser de 10 personas, ya que grupos con más participantes no obtenían mejores resultados en la resolución de problemas.

Otros autores como Surowiecki (2005), avalan la alta eficacia colectiva de grandes grupos en la clasificación y toma de decisiones, que en determinadas circunstancias pueden tener un acierto superior al de los expertos.

Según Surowiecki (2005), las decisiones colectivas tienen más probabilidades de ser buenas cuando las toman personas con opiniones distintas ya que elaboran conclusiones independientes, atendiendo primordialmente a la información privada de que disponen.

Por otra parte, el análisis sensorial es un método científico utilizado para obtener, medir, analizar e interpretar las reacciones frente a determinadas características de los alimentos y de los materiales, tal y como son percibidas por los sentidos de la vista, olfato, tacto, gusto y oído" (Stone y Sidel, 1993).

Aunque su aplicación principal es la del control de calidad, en función del protocolo que se siga puede tener múltiples utilidades.

Este análisis sensorial comprende un conjunto de técnicas para la medida precisa de las respuestas humanas a los alimentos, e intenta aislar las propiedades sensoriales aportando información útil para el desarrollo de productos, control durante elaboración, vigilancia durante el almacenamiento, predicción del éxito de un nuevo producto (Ramírez-Navas, 2012) e incluso, podrían utilizarse para la obtención de decisiones basadas en los principios de la inteligencia colectiva.

Algunas pruebas de análisis sensorial permiten traducir las sensaciones de los consumidores en atributos bien definidos para un producto y se les denomina pruebas orientadas al consumidor (Watts Ylimarki, Jeffery y Elias 1989).

Las pruebas de preferencia y aceptación son las más conocidas (Lawless y Heymann, 2010) y aunque se usan indistintamente son métodos distintos (Drake, 2007). En las pruebas de preferencia, a los consumidores se les solicita que indiquen cuál es la muestra preferida. Mientras que en las pruebas de aceptación a los consumidores se presentan productos y se les pide que indiquen el nivel de agrado en una escala. Las escalas se emplean con el objetivo de asegurar la validez de los métodos estadísticos utilizados en el procesado de datos (Torricella Moralles *et al.*, 2007).

Las pruebas de aceptación o aceptabilidad muestran la disposición de una persona a adquirir un producto (Watts *et al.*, 1989), y no solo dependen de la impresión que el juez reciba al probarlo, sino también de aspectos culturales, socioeconómicos y hábitos del consumidor (Anzaldúa-Morales, 1994).

Esta dependencia del juicio propio del consumidor permite que no sea necesaria la selección y entrenamiento de los jueces, ya que, en esta situación dicho entrenamiento pueden generar sesgos contraproducentes (Carpenter, Lyon y Hasdell, 2009). El hecho de prescindir de los jueces entrenados se traduce en una disminución de los costes de producción del análisis en tiempo y dinero.

Además de las pruebas de aceptabilidad y preferencia, dentro de las pruebas orientadas al consumidor, se encuentran las pruebas de valoración o “rating” y las pruebas de ordenación o “ranking”.

En el presente estudio se ha decidido emplear ambas pruebas ya que no existe una unanimidad a cerca de que prueba es más efectiva en el análisis sensorial, si una ordenación (Falahee y MacRae, 1997) o una valoración (Villanueva, Petenate y Da Silva, 2005).

Munson y McIntyre (1979) destacan que las escalas de valoración, son útiles para la evaluación de atributos por varios motivos; permiten que haya empates entre varias muestras, son fáciles de administrar y de completar y permiten el uso de métodos estadísticos paramétricos.

Las pruebas de ordenación se basan en la ordenación de un grupo de muestras en función de algún criterio específico. Los jueces tienen que elaborar un orden específico de todas las muestras comparadas unas con otras (Villanueva *et al.*, 2005). Una característica de los ranking es que se debe considerar una sucesión de productos cada cual más importante y sin permitir los empates.

Estos formatos, si se emplean de forma adecuada, permiten ordenaciones de atributos basadas en características que mediante otros métodos no serían evidentes. Además permiten que las pruebas contengan grandes cantidades de muestras que a su vez generan que el juez tenga que prestar mayor atención, dando lugar a datos más fiables (Alwin y Krosnick, 1985). Por otro lado, las pruebas de ordenación pueden generar una sobredimensión de las diferencias entre atributos que en ocasiones pueden no tener excesivo valor (Maio, Bell y Esses 1996).

2. Objetivo

El objetivo de este trabajo es analizar si las técnicas de inteligencia colectiva se pueden aplicar y son un método efectivo para sustituir a los sistemas clásicos de análisis sensorial, basado en el análisis estadístico de las valoraciones de los jueces.

Donde a partir de las opiniones de grupos numerosos de posibles consumidores sea posible la medición y/o predicción del éxito de un producto evaluando la reacción del mercado hacia el mismo.

3. Materiales y métodos

3.1. Muestras

El ensayo se realiza con dos series de muestras; una serie de 9 vinos tintos roble de diferentes D.O. españolas, y una serie de 9 galletas tipo cookie con chips de chocolate. En la Tabla 1 se muestran las galletas seleccionadas, mientras que los vinos seleccionados se pueden observar en la Tabla 2.

Tabla 1. Muestras de las galletas

	Nombre Comercial	Distribuidor
Muestra 1	Hacendado	Mercadona
Muestra 2	Chips Ahoy	Mondelez
Muestra 3	Carrefour	Carrefour
Muestra 4	Grandino	Lidl
Muestra 5	Chips Ahoy	Mondelez
Muestra 6	Alteza	Lupa
Muestra 7	American Cookies	Aldi
Muestra 8	DIA	DIA
Muestra 9	IFA Eliges	GADIS

Tabla 2. Muestras de los vinos

	Nombre Comercial	Denominación de Origen
Muestra 1	24 Mozas	D.O. Toro
Muestra 2	Viridiana	D.O. Ribera del Duero
Muestra 3	Elías Mora	D.O. Toro
Muestra 4	Cuatro Pasos	D.O. Bierzo

Muestra 5	Viridiana	D.O. Ribera del Duero
Muestra 6	Laya	D.O. Almansa
Muestra 7	La Planta	D.O. Ribera del Duero
Muestra 8	Juan Gil	D.O. Jumilla
Muestra 9	Honoro Vera	D.O. Calatayud

Se ha incorporado una muestra de referencia, en ambas series las muestras número 2 y 5 se corresponden con el mismo artículo.

Las muestras se presentaron utilizando un diseño de bloques completos, codificadas y en un orden aleatorio, de esta forma se evitan los posibles sesgos entre los jueces por la disposición relativa de las muestras, la sugestión que produce ver una marca conocida y las interacciones propias entre las muestras.

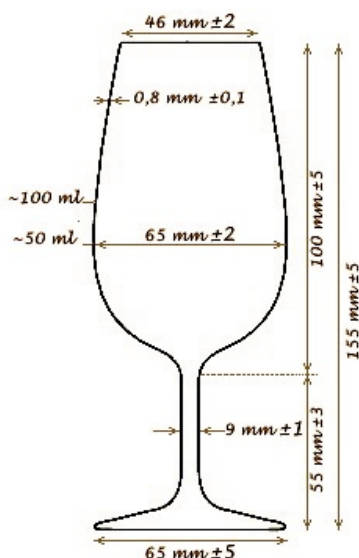


Figura 1. Catavinos UNE 87022:1992.

Las galletas se mostraron partidas a la mitad para evitar evidenciar que las diferencias de color, tamaño y número de pepitas de chocolate pudieran influir en las valoraciones de los jueces. El tamaño de la ración era de aproximadamente 15 g.

Los vinos se presentaron en copas de cata acreditadas (UNE 87022:1992), con una cantidad equivalente a 50 ml.

3.2. Pruebas

Se realizaron tres pruebas; aceptabilidad, ordenación y valoración; de acuerdo a la normativa UNE-ISO 6658:2008.

- Para la prueba de aceptabilidad se estableció una escala hedónica de 9 puntos en la que los jueces debían puntuar cada una de las muestras, (Watts *et al.*, 1989) (Figura 2.a).

- Para la prueba de ordenación se debían ordenar las 9 muestras en función de la impresión global, en orden descendente, de mayor a menor, (UNE-ISO 8587:2010) (Figura 2.b).
- Para la prueba de valoración se estableció un sistema de multivotación o multivoting, en el cual, el juez cuenta con 20 puntos a distribuir de la manera que creyera oportuno entre las 9 muestras, los empates estaban permitidos, (Klein y Bicharra, 2015) (Figura 2.c).

PRUEBA DE VALORACIÓN

Por último, valora las muestras, tienes **20 puntos** para repartir entre las **9 muestras**:

MUESTRA	PUNTOS
1	0
2	2
3	3 2
4	2
5	0
6	4
7	6
8	1
9	3

¡¡¡MUCHAS GRACIAS POR TU PARCIPACIÓN!!!

PRUEBA DE ACEPTABILIDAD

Evalúa cuánto te gusta o te disgusta la galleta señalando con un círculo la opción elegida:

GALLETA 1:

1	2	3	4	5	6	7	8	9
Me disgusta muchísimo	Me disgusta mucho	Me disgusta moderadamente	Me disgusta poco	Ni me gusta ni me disgusta	Me gusta poco	Me gusta moderadamente	Me gusta mucho	Me gusta muchísimo

PRUEBA DE ORDENACIÓN

A continuación **ordena las 9 muestras de mejor a peor**:

7	9	3	6	2	5	4	8	1	
				Mejor					Peor

Figura 2. a) Prueba de Aceptabilidad, b) Prueba de Ordenación y c) Prueba de Valoración.

3.3. Jueces

Para la realización del ensayo se ha contado con 252 participantes voluntarios, de los que el 50% eran hombres y el 50% mujeres, con un rango de edades de entre 16 y 24 años principalmente.

En la Tabla 3 se especifica el número de jueces que ha participado en cada prueba.

Tabla 3. Número de jueces y distribución

Prueba	Jueces cookies	Jueces vinos
Aceptabilidad	136	140
Ordenación	129	140
Valoración	20	75

Cabe destacar que, de acuerdo a la normativa UNE-ISO 8587:2010, para la determinación del orden de preferencia en un ensayo hedónico, el número mínimo de

jueces se determina en función del riesgo asumido, normalmente en torno a 60 jueces por grupo y tipo de consumidor como mínimo.

Para el análisis estadístico de los resultados, cuanto mayor es el número de jueces, mayor es la probabilidad de poner en evidencia cualquier diferencia sistemática en la ordenación entre productos.

3.4. Lugar de realización

Las pruebas de análisis sensorial se realizaron siguiendo la normativa UNE-ISO 6658:2008. Dichas pruebas se realizaron tanto en la sala de catas de la ETSIIAA del Campus de Palencia, como en diferentes institutos participantes. La duración de la sesión era variable, dependiendo del tiempo que necesitasen los jueces, pero no más de 35 minutos.

La sala de cata cuenta con 40 puestos o cubículos cumpliendo la normativa UNE-EN ISO 8589:2010/A12014 para análisis sensorial.

3.5. Análisis de datos

3.5.1. Análisis de datos desde el punto de vista sensorial

El análisis de los datos obtenidos en la prueba de aceptabilidad se realizó mediante una prueba de análisis de varianza o ANOVA, factor muestras (Villanueva, *et al.*, 2005). Posteriormente, se aplicó una prueba post hoc de comparación de medias, test de Tukey, para comprobar entre qué muestras existe una diferencia estadísticamente significativa.

El ANOVA se trata de una prueba paramétrica que permite la comparación entre dos o más poblaciones de distribución normal, a partir de las variaciones en las medias del factor estudiado.

Para el análisis de datos de la prueba de ordenación se ha recurrido tanto a la normativa UNE-ISO 8587:2010 como a Lehmann (1975), Lawles y Heymann (2010) y Rodríguez *et al.*, (2000). Primero se debe realizar un test de Friedman para valorar si existen diferencias significativas entre las ordenaciones de los productos, y en caso de que las hubiera se calcularía la Mínima Diferencia Significativa (MDS) para el riesgo asumido $\alpha=0,05$, con la que se valorará qué productos son significativamente diferentes entre sí. Se trata de una prueba desarrollada también por Friedman que compara los distintos valores del factor con un crítico. Permite obtener diferencias significativas reales aún cuando hay pocas observaciones en cada tratamiento

El Test de Friedman es una prueba no paramétrica desarrollado por Milton Friedman equivalente a la prueba ANOVA para medidas repetidas en la versión no paramétrica.

La prueba de valoración ha requerido un análisis previo para conocer la distribución de los datos, test de Kolmogorov-Smirnov. Posteriormente se ha aplicado un análisis de varianza para poblaciones no paramétricas o H de Kurskal-Wallis (Sayadi, Erraach, Parra-López y Jiménez-López, 2013). La prueba H es una extensión de la prueba de la U de Mann-Whitney para 3 o más grupos.

El objeto de esta prueba es indicar si existe o no variabilidad entre las valoraciones, lo que se traduce en la existencia de diferencias entre las muestras examinadas. Para finalizar, en caso de que se hayan detectado diferencias, se realizará una prueba post hoc de comparación de medianas.

El tratamiento estadístico de los datos se ha realizado con el programa informático IBM SPSS Statistics V23.0 para Windows.

3.5.2. Análisis de datos desde el punto de vista de la inteligencia colectiva

A diferencia del tratamiento de los datos desde el punto de vista del análisis sensorial, basado en técnicas estadísticas, el tratamiento de los datos desde el punto de vista de la inteligencia colectiva se basa en agregaciones de resultados obtenidos a partir de las opiniones de los jueces.

Las pruebas de aceptabilidad y valoración siguen el mismo método de agregación. En ambas se realizan sumatorios del total de las puntuaciones asignadas por los jueces a cada muestra. Posteriormente se comparan los resultados en parejas y se asigna una puntuación a cada muestra. De esta forma se obtiene un nuevo orden o ranking agregado a partir de las puntuaciones obtenidas en las comparaciones.

Por otro lado, para la prueba de ordenación el método empleado se basa en la aplicación del algoritmo MedRank (Fagin, Kumar y Sivakumar, 2003). Este algoritmo se encarga de la elaboración de un nuevo ranking basado en el orden de aparición de las muestras en al menos la mitad de los jueces participantes. De tal manera que en dicho ranking el primer puesto lo ocupará aquella muestra que aparezca en primera posición en más de la mitad de los juicios emitidos, y así sucesivamente.

4. Resultados y discusión

4.1. Galletas

Los resultados del ANOVA obtenidos con los datos de la prueba de aceptabilidad dan una $F=8,096$ con un $p\text{-valor}=0,000$, por lo tanto se puede decir que existen diferencias estadísticamente significativas entre la aceptación de las distintas muestras de galletas. En la Tabla 4 se observan los resultados del test de Tukey, dicha prueba tiene como finalidad conocer qué muestras son significativamente diferentes.

Tabla 4. Valores medios de aceptabilidad galletas

Muestras	Aceptabilidad
1	6,34 ± 2,04 ^{bc}
2	6,08 ± 1,67 ^b
3	6,56 ± 1,80 ^{bc}
4	6,01 ± 2,23 ^{ab}
5	5,96 ± 1,80 ^{ab}
6	5,88 ± 1,58 ^{ab}
7	6,94 ± 1,50 ^c
8	5,37 ± 2,17 ^{ab}
9	5,86 ± 1,81 ^{ab}

Los superíndices con letras distintas indican diferencias estadísticas significativas entre las muestras después de la aplicación del test de Tukey.

La tabla muestra que existen 3 subconjuntos diferentes. Cada subconjunto engloba las muestras que no presentan diferencias significativamente entre sí.

Cabe destacar que la desviación estándar para las muestras es considerablemente pequeña para el número de datos manejados.

De acuerdo a la puntuación obtenida en función de la prueba de aceptabilidad, las muestras se ordenan de la siguiente manera:

7 American Cookies > 3 Carrefour > 1 Hacendado > 2 Chips Ahoy > 4 Grandino > 5 Chips Ahoy > 6 Alteza > 9 IFA Eliges > 8 DIA.

Los resultados obtenidos para la prueba de ordenación mediante el test de Friedman, dan un F-valor de 59,339 y un p-valor inferior a 0,001. Dado que el error considerado era de $\alpha = 0,05$ se puede afirmar que existen diferencias significativas entre las ordenaciones de galletas.

Los resultados de las pruebas post hoc se representan gráficamente agrupando las distintas muestras en función de sus diferencias significativas. De forma que, los productos subrayados por una misma línea se interpreta como que no presentan diferencias estadísticamente significativas entre sus ordenaciones. Mientras que aquellos productos no subrayados por una misma línea continua se interpreta que han obtenido ordenaciones significativamente diferentes para el riesgo asumido.

En la Figura 3 se pueden observar cuatro grupos de muestras atendiendo a los diferentes niveles de significación.

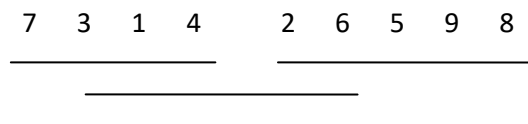


Figura 3. Representación Ordenación galletas

El primer grupo incluye a las muestras 7 American Cookies, 3 Carrefour, 1 Hacendado, y 4 Grandino, el segundo incluye a las muestras 3, 1, 4, 2 y 6; el tercer grupo está formado por las muestras 1, 4, 2, 6, 5 y 9; y por último el grupo compuesto por las muestras 2 Chips Ahoy, 6 Alteza, 5 Chips Ahoy, 9 IFA Eliges y 8 DIA,

De esta manera, obtenemos el siguiente orden:

7 American Cookies > 3 Carrefour > 1 Hacendado > 4 Grandino > 2 Chips Ahoy > 6 Alteza > 5 Chips Ahoy > 9 IFA Eliges > 8 DIA.

Los resultados obtenidos consideran a la muestra 7 American Cookies como la muestra mejor ordenada estadísticamente, y a la muestra 8 DIA como la peor ordenada estadísticamente.

A su vez, los jueces han ordenado a las muestras 2 y 5, ambas correspondientes con la marca Chips Ahoy, a muy poca distancia la una de la otra, sin que presenten diferencias significativas entre sí.

En cuanto al análisis de los datos obtenidos en la prueba de valoración, el primer paso es comprobar la distribución de los datos obtenidos. Dado el tamaño de la serie de datos se aplicará la prueba de Kolmogorov-Smirnov. El resultado nos da un p-valor inferior $< 0,001$, por lo que se trata de una distribución no paramétrica.

Tabla 5. Prueba de normalidad galletas

Kolmogorov-Smirnov			
	Estadístico	gl	Sig.
Puntuación	0,217	153	0,000

La prueba de elección para distribuciones no paramétricas es la H de Kruskal-Wallis. Esta prueba da un F-valor de 16,584 y un P-valor de 0,035, ligeramente inferior al error asumido $\alpha = 0,05$. Este resultado rechaza la hipótesis nula y confirma la existencia de diferencias significativas entre las muestras examinadas. Sin embargo, después de aplicar la comparación de medianas para ver en qué muestras existen diferencias significativas, se observa que no existen diferencias entre ninguna de las muestras.

Se puede afirmar que el orden de la valoración es el siguiente a pesar de que en la comparación de medianas no se hayan observado diferencias significativas entre las muestras de galletas:

9 IFA Eliges > 7 American Cookies > 2 Chips Ahoy > 3 Carrefour > 4 Grandino > 1 Hacendado > 6 Alteza > 8 DIA > 5 Chips Ahoy.

Se observan variaciones respecto a los resultados obtenidos a partir de las pruebas anteriores. En esta situación las muestras 2 y 5 correspondientes con Chips Ahoy se encuentran separadas por una amplia distancia. Aún así, en esta prueba, la muestra 7 American Cookies aparece como una de las mejor valoradas y la muestra 8 DIA aparece como de las peores.

En la Tabla 6 aparecen resumidos los resultados de las tres pruebas, después del análisis de datos desde el punto de vista sensorial.

Tabla 6. Resultado análisis estadístico

Prueba	Resultados
Aceptabilidad	7> 3> 1> 2> 4 > 5> 6> 9> 8
Ordenación	7> 3> 1> 4> 2> 6> 5> 9> 8
Valoración	9> 7> 2> 3> 4> 1> 6> 8> 5

A continuación, los datos de las tres pruebas; aceptabilidad, ordenación y valoración; se han sometido a técnicas de agregación de resultados, propias del análisis de la inteligencia colectiva. El objetivo es obtener nuevos rankings, desde el punto de vista

de la IC, y poder comparar los resultados con los obtenidos mediante el análisis estadístico.

La prueba de aceptabilidad ha obtenido el siguiente orden agregado:

7 American Cookies > 3 Carrefour > 1 Hacendado > 2 Chips Ahoy > 4 Grandino > 5 Chips Ahoy > 6 Alteza > 9 IFA Eliges > 8 DIA.

Los resultados del tratamiento de datos con el algoritmo MedRank para la prueba de ordenación son:

7 American Cookies > 3 Carrefour > 4 Grandino > 1 Hacendado > 6 Alteza > 5 Chips Ahoy > 9 IFA Eliges > 2 Chips Ahoy > 8 DIA.

El nuevo ranking agregado resultante de las pruebas de valoración obtenido es el siguiente:

9 IFA Eliges > 7 American Cookies > 2 Chips Ahoy > 3 Carrefour > 4 Grandino > 1 Hacendado > 6 Alteza > 8 DIA > 5 Chips Ahoy.

En la Tabla 7 aparecen reflejados los tres rankings.

Tabla 7. Resultados análisis IC galletas

Prueba	Resultados
Aceptabilidad	7> 3> 1> 2> 4 > 5> 6> 9> 8
Ordenación	7> 3> 4> 1> 6> 5> 9> 2> 8
Valoración	9> 7> 2> 3> 4> 1> 6> 8> 5

Si se comparan los resultados obtenidos después del análisis de datos tanto desde el punto de vista sensorial (Tabla 6) como desde el punto de vista de la inteligencia colectiva (Tabla 7), se observan los siguientes resultados. En la Tabla 6, se pueden apreciar ciertas similitudes, principalmente entre el ranking obtenido a partir de la prueba de aceptabilidad y el ranking obtenido a partir de la prueba de ordenación. En ambos rankings tanto la galleta 7 (American Cookies) como la galleta 3 (Carrefour) se consideran las mejores y la galleta 8 (DIA) es la considerada peor en ambos rankings.

Atendiendo a las galletas situadas en posiciones intermedias, se puede observar que no hay una variación excesiva entre ambas pruebas, salvo la galleta 2 (Chips Ahoy) que pasa de la cuarta posición en el ranking de aceptabilidad a la penúltima en el ranking de ordenación. Aún así, en ambas pruebas ha sido igualmente ordenada cerca de la muestra 5 (Chips Ahoy).

Los resultados del ranking de valoración, muestran un cierto nivel de discordancia con las pruebas anteriores, ya que la galleta 9 (IFA Eliges) ha sido ordenada en primera posición, la galleta 7 (American Cookies) ha sido ordenada en segunda posición y las galletas 8 (DIA) y 5 (Chips Ahoy) en penúltima y última posición respectivamente. Además, existe una diferencia entre la ordenación de las muestras de referencia correspondientes a la marca Chips Ahoy (galletas 2 y 5).

A pesar de estas ligeras diferencias, se puede considerar que, en general, existe cierto acuerdo en considerar a la galleta 7 (American Cookies) como la mejor y a la galleta 8 (DIA) como la peor.

La Tabla 7 muestra los resultados del análisis desde el punto de vista de la inteligencia colectiva. Las conclusiones son las mismas que las obtenidas analizando la Tabla 6. Las tres pruebas ordenan a la muestra 7 American Cookies entre las mejores galletas y a la muestra 8 DIA entre las peores. El resto de muestras han sido ordenadas en posiciones similares pero no idénticas entre sí.

Además, salvo en la prueba de valoración, las muestras 2 y 5 se han ordenado muy próximas entre sí, y en varias pruebas han sido situadas entre los últimos puestos.

Siguiendo el hilo de las comparaciones, si confrontamos los rankings obtenidos a partir del tratamiento estadístico, Tabla 6, con los rankings obtenidos a partir de técnicas de agregación, Tabla 7, se evidencian numerosas coincidencias entre ambos métodos.

El ranking obtenido en sendas pruebas de aceptabilidad coincide, independientemente del método de obtención y es el siguiente:

7 American Cookies > 3 Carrefour > 1 Hacendado > 2 Chips Ahoy > 4 Grandino > 5 Chips Ahoy > 6 Alteza > 9 IFA Eliges > 8 DIA.

Al igual que con las pruebas de aceptabilidad, las pruebas de valoración obtenidas a partir del análisis estadístico y las obtenidas a partir de los resultados de agregación de los resultados también coinciden.

9 IFA Eliges > 7 American Cookies > 2 Chips Ahoy > 3 Carrefour > 4 Grandino > 1 Hacendado > 6 Alteza > 8 DIA > 5 Chips Ahoy.

Las pruebas de ordenación, por otro lado, presentan variaciones en la ordenación de las muestras, dependiendo de si ésta proviene del análisis sensorial o del análisis algorítmico. Dichas variaciones alteran la posición de algunas muestras uno o dos puestos. Y aunque existan similitudes no son tan evidentes como en las pruebas anteriores.

Pese a las mencionadas diferencias, los puntos en común entre ambos métodos son fundamentales. Principalmente, tanto la muestra 7 (American Cookies) como la muestra 8 (DIA) han sido seleccionadas como mejor y peor ordenadas. Las muestras de referencia 2 y 5 correspondientes con Chips Ahoy se encuentran ordenadas a poca distancia en ambos métodos.

A pesar de la diferencia entre el número de participantes de cada una de las pruebas de aceptación, ordenación y valoración, los resultados son bastante concluyentes.

Con estos datos, se puede afirmar que existen evidencias para equiparar el análisis mediante técnicas de inteligencia colectiva al análisis sensorial, de forma que, aunque no exista una exactitud entre los resultados de uno y otro método, ambos métodos presentan un consenso a la hora de escoger los productos mejores y peores.

Para conocer el origen de estas diferencias en las ordenaciones de las muestras se ha recurrido al estudio de la consultora Contract Testing "Are there real differences

between brands” (2014), en el que se analizaban las diferencias entre distintas marcas de galletas tipo cookies.

Dicho estudio señala al sabor, y particularmente al sabor a chocolate, como la característica más influyente en la toma de decisiones por parte del consumidor. También destaca la importancia del aspecto visual ya que complementa al sabor. Cuanto más visibles sean las pepitas de chocolate, mayor percepción de sabor a chocolate habrá por parte de los consumidores.

Otros atributos influyentes muy valorados en la elección de una galleta tipo cookie son el tamaño de los chips de chocolate y el regusto que deja la galleta.

De la misma forma, las características negativas señaladas en el estudio son el ratio chips de chocolate/galleta y la intensidad del azúcar. Cuanto menor fueran ambos valores peores resultados de aceptabilidad de las galletas.

La textura de las galletas se considera un factor contributivo pero no actúa de forma autónoma, es decir, es un factor importante pero no actúa de manera independiente.

Para realizar una comparación más visual se ha elaborado una tabla resumen, Tabla 8, con los principales datos de las galletas testadas en este trabajo.

Los resultados del análisis estadístico coinciden con los resultados del análisis de agregación en señalar a la muestra 7 American Cookies como la muestra mejor ordenada. Otra muestra que sale muy bien parada de la comparación es la muestra 3 Carrefour, ya que se ha ordenado como la segunda mejor en ambos análisis.

Ambas galletas se caracterizan por presentar niveles de azúcar del 37%. El porcentaje de chocolate presenta diferencias entre ambas galletas, la muestra 7 American Cookies, presenta un porcentaje superior de chocolate 40% total, dividido en 29% chocolate negro y 11% chocolate con leche. Respecto de la muestra 3 Carrefour presenta un 37% sin especificar si es chocolate negro o con leche.

Tabla 8. Comparativa galletas

Muestra	Marca	Supermercado	Fabricante /importador	Chocolate	Azúcar
1	Hacendado	Mercadona	Siro	37%	39%
2/5	Chips Ahoy	Gadis, Día, Carrefour	Nabisco	25,6 %	64%
3	Carrefour	Carrefour	Países Bajos para ARLUY S.L.U.	37%	37 %
4	Grandino,	Lidl	Galletas Gullón	29% Negro + 11 Con leche	36%
6	Alteza s	Lupa	Galletas Gullón	25%	31%

7	American Cookies	Aldi	Banketakkerij Merba B.V	29% chocolate y 11% chocolate con leche	37,3 %
8	Cookies DIA	DIA	Dan Cake Portugal	29% Negro + 11% Con leche	40%
9	IFA Eliges	Gadis	Galletas Gullón	25%	31%

Tanto el análisis estadístico como el de I.C también han coincidido al señalar a la galleta 8 DIA como la peor ordenada. Otras muestras con ordenaciones negativas han sido las muestras correspondientes con las galletas IFA Eliges y Chips Ahoy.

Estas muestras se han caracterizado por presentar un porcentaje de azúcar superior a los de cacao. Destacan las galletas Chips Ahoy que con un 64% de azúcar doblan al porcentaje de cacao, 25,6%. A pesar de que los valores de cacao, 25%, de la muestra IFA Eliges también son inferiores a los de azúcar, 31%, su diferencia no es tan abultada como en la muestra anterior.

Un caso especial es la muestra 8 de DIA, pues si solo se atienden a estos indicadores, azúcar 40%, y cacao 40% (29% Chocolate Negro y 11% Chocolate con leche), no difieren demasiado de los de la galleta mejor valorada. Por lo que induce a pensar que existan otras variables que sin haberse tenido en cuenta, pueden influenciar en las opiniones de los jueces.

Estudios anteriormente publicados (Jones, 2003; Colleen, 2007; Doolittle, 2007) apuntan a la grasa como una de esas posibles variables a tener en cuenta, ya que está contribuye al olor, a la textura y a la aceptabilidad final.

Autores como Cauvain y Young (2006) destacan por su parte, la influencia de la receta en el proceso de elaboración del producto, ya que aún empleando los mismos ingredientes, diferencias en las cantidades o calidades de los mismos pueden generar diferencias de tamaño, forma y textura muy relevantes en el producto final.

A pesar de que algunas de las muestras analizadas presentaban grasas, derivados del huevo y de la leche, para favorecer la palatabilidad y mejorar el aspecto externo del producto, no se ha podido confirmar que ninguna de estas variables haya influido en el resultado de las pruebas de aceptabilidad, ordenación y valoración. De la misma forma, se carecen de datos que aporten información acerca del ratio chocolate/galleta o la intensidad del azúcar empleado en las muestras testadas.

Por lo que para explicar las diferencias presentes entre las ordenaciones de las muestras, se ha supuesto que estas diferencias atienden a las diferencias presentes entre la calidad de los ingredientes y las recetas empleadas.

En conclusión, no se puede afirmar que en el presente estudio la concentración del chocolate haya sido relevante para los jueces a la hora de establecer un orden, sin embargo y a falta de mayor información al respecto, parece que la mayor presencia de azúcar si puede haber influido en el posicionamiento final de las muestras.

4.2. Vinos

Los resultados del ANOVA obtenidos con los datos de la prueba de aceptabilidad dan una $F=4,74$ con un p-valor inferior a $\alpha=0,05$, por lo tanto se puede afirmar la existencia de diferencias estadísticamente significativas para las ordenaciones de la prueba de aceptación. En la Tabla 9 se observan los resultados del test de Tukey para las muestras de vino.

Tabla 9. Valores medios de aceptabilidad de vinos

Muestras	Aceptabilidad
1	4,93 ± 2,14 ^{abcd}
2	5,60 ± 1,96 ^d
3	4,79 ± 1,99 ^{abc}
4	4,42 ± 2,14 ^a
5	5,35 ± 1,81 ^{bcd}
6	4,96 ± 2,10 ^{abcd}
7	5,53 ± 2,23 ^{cd}
8	4,72 ± 2,09 ^{ab}
9	5,13 ± 2,08 ^{abcd}

Los superíndices con letras distintas indican diferencias estadísticas significativas entre las muestras después de aplicar el test de Tukey.

Los resultados de la Tabla 9 muestran 4 subconjuntos. Cada subconjunto incluye muestras con diferencias estadísticamente no significativas entre sí.

Al igual que con la serie de galletas anterior, la desviación estándar de las muestras evaluadas es considerablemente pequeña para el tamaño de la población estudiada.

De esta forma, finalizado el análisis estadístico, se obtiene el siguiente orden para la prueba de aceptabilidad:

2 Viridiana > **7** La Planta > **5** Viridiana > **9** Honoro Vera > **6** Laya > **1** 24 Mozas > **3** Elías Mora > **8** Juan Gil > **4** Cuatro Pasos.

Los resultados de la prueba de Friedman para la prueba de ordenación dan un valor del estadístico F de 54,671, con un nivel de significación, p-valor inferior a 0,001, lo que rechaza la hipótesis inicial y confirma que existen diferencias significativas entre las ordenaciones para un nivel de significación $\alpha= 0,05$.

En la Figura 4, se observan los resultados de las prueba post hoc. Las muestras se agrupan en función de las diferencias significativas que presentan sus ordenaciones. De manera que, los productos subrayados por la misma línea se pueden interpretar como que no presentan diferencias significativas entre sus ordenaciones, mientras que los no subrayados por la misma línea si presentan dichas diferencias entre sus ordenaciones para el nivel de riesgo asumido.

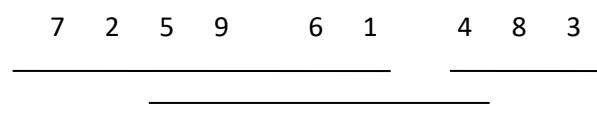


Figura 4. Representación ordenación vinos

El resultado se ordena de la siguiente forma:

7 La Planta > 2 Viridiana > 5 Viridiana > 9 Honoro Vera > 6 Laya > 1 24 Mozas > 4 Cuatro pasos > 8 Juan Gil > 3 Elías mora

Analizando los resultados se observan 5 grupos. Un primer grupo compuesto por las muestras 7, 2, 5, 9, 6 y 1 (La Planta, Viridiana, Viridiana, Honoro Vera, Laya y 24 Mozas), un segundo grupo en el que se encuentran los vinos 4, 8 y 3 (Cuatro Pasos, Juan Gil y Elías Mora), un tercer grupo formado por los vinos 2, 5, 9, 6, 1 y 4 (Viridiana, Viridiana, Honoro Vera, Laya, 24 Mozas y Cuatro Pasos), un cuarto grupo compuesto por las muestras 9, 6, 1, 4, 8 y 3 (Honoro Vera, Laya, 24 Mozas, Cuatro Pasos, Juan Gil y Elías Mora), y finalmente el quinto grupo, compuesto por los vinos 6, 1, 4, 8 y 3 (Laya, 24 Mozas, Cuatro Pasos, Juan Gil).

Al igual que con las galletas, las muestras 2 y 5 corresponden a la misma referencia, Viridiana, y la ordenación de ambos artículos puede dar información muy relevante a la hora de comparar los métodos.

La prueba de valoración comienza con un estudio de la distribución de los datos mediante la prueba de Kolmogorov. Los resultados indican un p-valor inferior a 0,000, por lo que se trata de una prueba no paramétrica.

Tabla 10 Prueba de normalidad vinos.

Kolmogorov-Smirnov			
	Estadístico	gl	Sig.
Puntuación	0,215	558	0,000

Los resultados de la prueba H dan un F-valor de 32,603, con una significación inferior a 0,001. Se confirma que existen diferencias significativas entre las valoraciones de las muestras y se rechaza la hipótesis nula.

La aplicación de las pruebas post hoc de comparación de medianas permite identificar cual de las muestras presenta diferencias significativas. Ésta revela que solo existen diferencias significativas entre la muestra 2 (Viridiana) y las muestras 3 y 8 (Elías Mora y Juan Gil), el resto de muestras no presentan ningún tipo de diferencia significativa.

Con estos resultados se puede afirmar que el orden de valoración es el siguiente:

2 Viridiana > 7 La Planta > 1 24 Mozas > 9 Honoro Vera > 5 Viridiana > 4 Cuatro Pasos > 6 Laya > 8 Juan Gil > 3 Elías Mora.

Tabla 11. Resultados análisis estadístico vino

Prueba	Resultados
Acceptabilidad	2 > 7 > 5 > 9 > 6 > 1 > 3 > 8 > 4.

Ordenación	7 > 2 > 5 > 9 > 6 > 1 > 4 > 8 > 3
Valoración	2 > 7 > 1 > 9 > 5 > 4 > 6 > 8 > 3.

Los resultados estadísticos de las pruebas de aceptabilidad, ordenación y valoración se encuentran en la Tabla 11. La comparación de los tres métodos coincide al situar en las primeras posiciones a las muestras 2 y 7 (Viridiana y La Planta). También coincide al ordenar los vinos 3 y 8 (Elías Mora y Juan Gil) entre las últimas posiciones.

A pesar de las semejanzas que guardan las tres pruebas, presentan algunas diferencias muy significativas. La más relevante está referida a las posiciones de las muestras 2 y 5 (Viridiana), las cuales, en las prueba de aceptabilidad y ordenación se han situado muy cerca la una de la otra, mientras que, en la prueba de valoración, se han ordenado a gran distancia. Por lo demás, existe un parecido razonable entre las ordenaciones de las tres pruebas estadísticas.

Los resultados de las pruebas sensoriales de aceptabilidad, ordenación y valoración se han sometido a las técnicas de agregación de IC para obtener nuevos rankings que permitan una mejor comparación intra y entre análisis.

Los resultados del análisis de I.C para la prueba de aceptabilidad son:

2 Viridiana > **7** La planta > **5** Viridiana > **9** Honoro Vera > **6** Laya > **1** 24 Mozas > **3** Elías Mora > **8** Juan Gil > **4** Cuatro Pasos.

La prueba de ordenación obtuvo los siguientes resultados:

7 La planta > **5** Viridiana > **2** Viridiana > **9** Honoro Vera > **6** Laya > **1** 24 Mozas > **4** Cuatro Pasos > **8** Juan Gil > **3** Elías Mora.

El nuevo ranking de valoración es el siguiente:

2 Viridiana > **7** La planta > **9** Honoro Vera > **1** 24 Mozas > **4** Cuatro Pasos > **5** Viridiana > **6** Laya > **3** Elías Mora > **8** Juan Gil.

Tabla 12. Resultados análisis IC vino

Prueba	Resultados
Ranking Aceptabilidad	2 > 7 > 5 > 9 > 6 > 1 > 3 > 8 > 4.
Ranking Ordenación	7 > 5 > 2 > 9 > 6 > 1 > 4 > 8 > 3
Ranking Valoración	2 > 7 > 9 > 1 > 4 > 5 > 6 > 3 > 8.

Como se puede apreciar en la Tabla 12, en general los resultados de las pruebas de aceptabilidad y de ordenación son muy similares entre sí, y como ocurría con las muestras de galletas, los resultados de la prueba de valoración son los que presentan mayores diferencias.

Tanto en la prueba de aceptabilidad, como en las de ordenación y valoración, las muestras 7 y 2 (La Planta y Viridiana) vuelven a estar entre las primeras posiciones, y las muestras 8 y 3 (Juan Gil y Elías Mora) entre las últimas.

Adicionalmente, en dos de las pruebas, aceptabilidad y ordenación, las muestras 2 y 5, correspondientes con el vino Viridiana, vuelven a situarse muy próximas en el ranking.

La comparación de los resultados del análisis estadístico, Tabla 11, con los del análisis de la inteligencia colectiva, Tabla 12, presenta bastantes puntos en común.

Para empezar, los resultados de las pruebas de aceptabilidad son idénticos para ambos análisis. La muestra 2 (Viridiana) es la mejor ordenada, muy próxima a la muestra 5, la otra muestra correspondiente con la referencia Viridiana. Asimismo, el vino 4 (Cuatro Pasos) es el peor ordenado.

La principal diferencia que presentan las pruebas de ordenación es que las muestras 2 y 5 correspondientes con la referencia Viridiana, han intercambiado sus posiciones.

Por lo general, ambas coinciden tanto en la mejor ordenada, muestra 7 La planta, como en la peor ordenada, muestra 3 Elías Mora.

No obstante, es en la prueba de valoración donde se encuentran más diferencias, y aún así, ambos análisis prácticamente coinciden en sus ordenaciones. Las muestras mejor ordenadas vuelven a ser la 2 y la 7 (Viridiana y La Planta) y las peores las muestras 8 y 3 (Juan Gil y Elías Mora). Las muestras 2 y 5 (Viridiana) en este caso se encuentran ordenadas a gran distancia.

Existen pequeñas variaciones entre las muestras ordenadas en posiciones intermedias, las cuales aparecen intercambiadas pero, en un orden muy similar al obtenido mediante análisis estadístico.

El vino es una matriz compleja con una gran variedad de componentes químicos; volátiles y no volátiles que generan diferencias en el sabor y aroma.

La evaluación de un tinto se basa principalmente en el análisis sensorial. Los análisis químicos se realizan para explicar posibles cambios observados en las pruebas sensoriales. La relación entre la evaluación sensorial y la composición química es un tema crítico en la investigación enológica, (Girard et al., 2001).

La combinación de la química del sabor con las técnicas de análisis sensorial ha permitido asociar las propiedades del vino a compuestos químicos específicos. Dentro de los compuestos asociados a propiedades organolépticas encontramos a los compuestos fenólicos.

Estos compuestos incluyen a los antocianos y proantocianos; relacionados con el color del vino y su estabilidad; o los taninos condensados; más vinculados con la estabilidad coloidal, el amargor y la astringencia del vino.

De acuerdo con Sáenz-Navajas (2015), existe un desajuste entre la percepción de un vino por parte de los profesionales del vino y de los consumidores, ya que, si los juicios de los expertos suelen basarse en procesos técnicos de elaboración del vino, las evaluaciones de los jueces no entrenados suelen basarse en experiencias individuales y subjetivas.

De esta manera, se concluye en que la percepción final de los vinos es muy variable y dependen de lo que se ha clasificado de forma general como atributos o factores intrínsecos y extrínsecos (Charters y Pettigrew, 2007):

- Los atributos intrínsecos se definen como “drinking experience” y engloban el aroma, sabor, sensación en boca y placer. Son los atributos más importantes ya que pueden influir de gran manera en la percepción y elección de un vino (Hopfer y Heymann, 2014; Pelsmaeker *et al.*, 2017).
- Los factores extrínsecos incluyen todos aquellos relacionados con el proceso de crecimiento de la uva y la elaboración del vino así como de todo el proceso relacionado con la venta y comercialización del producto, Casini, Corsi y Goodman (2009).

La importancia de los atributos intrínsecos ya fue identificada en estudios previos como los de Nunes, Madureira, Oliveira y Madureira, (2016) o Goodman, (2009) y recaen en la experiencia previa como factor determinante para la elección de un vino. Dichos estudios señalan que además de factores intrínsecos existe cierta influencia de ciertos factores extrínsecos; como recomendaciones por parte de amigos o la región de origen.

La importancia de estos atributos extrínsecos recae en que el consumidor, por norma general, no tiene opción de probar el producto previamente a su elección y por lo tanto, depende de factores externos (Sáenz-Navajas *et al.*, 2016).

Para facilitar la comparación entre las muestras de vino, se ha realizado una tabla resumen, Tabla 13, con los principales datos obtenidos a partir de las muestras de vino empleadas.

Dado que la comparación de las pruebas estadísticas y las pruebas de agregación presentan resultados similares, se procede a un análisis conjunto de los resultados.

Tabla 13. Tabla comparativa vinos

Muestra	Nombre	Meses en barrica	Graduación	Uva	Bodega	D.O
1	24 Mozas	6	14,5%	Tinta de Toro	Divina Proporción	Toro
2 – 5	Viridiana	3	13,5%	Tempranillo	Atalaya de Golbán	Ribera del Duero
3	Viña Elías Mora	6	14,5%	Tinta de Toro	Elias Mora	Toro
4	Cuatro pasos	2-3	13,5%	Mencia	Cuatro Pasos	Bierzo
6	Laya	6	14%	Garnacha tintorera + Monastrell	Atalaya GFE	Almansa
7	La Planta	6	14,5%	Tempranillo	Arzuaga Navarro	Ribera del Duero
8	Juan Gil Amarillo	-	15%	Monastrell	Juan Gil GFE	Jumilla
9	Honoro Vera	2	14%	Garnacha	Ateca GFE	Calatayud

Ambos métodos, estadístico e IC, coinciden en ordenar a las muestras 2 Viridiana en primera posición seguida de cerca por la muestra 7 La Planta. Las muestras peor ordenadas corresponden con los vinos 8 Juan Gil y 3 Elías Mora, aunque atendiendo a

la prueba de aceptabilidad se observa que la muestra 4 Cuatro Pasos presenta menor aceptación que estas.

Como ya se ha mencionado, existen diferencias entre las opiniones de los jueces en función del grado de especialización de los mismos. Se suele afirmar que los jueces inexpertos se basan principalmente en experiencias previas individuales y subjetivas.

Con los datos de los que se dispone no se puede evidenciar que haya una correlación entre los resultados y la graduación alcohólica o la edad de maduración del vino, pero sí existe una relación entre la posición y el tipo de uva.

La mayoría de los vinos están elaborados a partir de una única variedad de uva, y los resultados, permiten apreciar cierto patrón en la distribución de las posiciones. De esta forma, las primeras posiciones son ocupadas por las uvas Tempranillo, mientras, la uva Garnacha presenta posiciones intermedias. Y en últimas posiciones encontramos la Mencía, Monastrell y Tinta de Toro.

La uva Tempranillo está considerada como una de las variedades uvas más apreciadas del mundo y es muy valorada organolépticamente de acuerdo con el Consejo Regulador D.O.Ca Rioja. Es capaz de producir vinos con largo envejecimiento, muy equilibrados en grado alcohólico, color y acidez, y con un paladar suave y afrutado, que evoluciona a tonos aterciopelados cuando envejece.

La uva Garnacha es una uva originaria de España pero muy extendida entre los países productores. Se le considera doble aptitud, para vino y para mesa. Produce vinos aromáticos, con alto contenido alcohólico, cuerpo medio y buena acidez, Sin embargo, se caracteriza por su rápido envejecimiento y su sensibilidad a la oxidación.

La Garnacha tintorera es una variedad derivada de la Garnacha que da lugar a vinos de gran intensidad, con mucho color, extracto seco y acidez, lo que le hace excelente para dar estructura en mezclas.

La uva Mencía posee color intenso frambuesa con tonos florales. El aroma posee notas de flores, con frutas rojas y negras silvestres. En boca es aterciopelada, su buen grado en acidez hace vinos muy dulces.

La uva Monastrell se caracteriza por ser muy dulce con sabor aromático y frutal. Produce vinos con una alta graduación alcohólica y un elevado nivel de taninos. Los vinos destacan por ser potentes en su juventud pero mejoran tras una mezcla o con una crianza. Destaca su uso para la elaboración de vinos dulces de postre.

La Tinta de Toro se considera una variedad específica de la uva Tempranillo, genera vinos equilibrados, alcohólicos, potentes, ligeramente ácidos, de color rubí intenso y con gran resistencia a la oxidación.

Existen diferencias entre las ordenaciones de los vinos de uva Tempranillo y Garnacha 100%, situados en las cuatro primeras posiciones, con los vinos situados en las tres últimas posiciones, Mencía, Monastrell y Tinta de toro.

Respecto a las diferencias entre los resultados de los vinos de uva Tempranillo y los vinos de Tinta de Toro, se ha demostrado que en vinos elaborados a partir del mismo

tipo de uva, tanto las levaduras empleadas como las técnicas utilizadas, aportaban propiedades organolépticas características y diferenciables, que podían medirse fácilmente mediante análisis fisicoquímicos (Girad *et al.*, 2011). De esta forma, se explica que los resultados hayan sido tan diferentes teniendo en cuenta que se estaban empleando las mismas materias primas.

En relación con el caso del vino de Juan Gil, la explicación más probable puede deberse a que la uva Monastrell genera un tipo de vino dulce, aromático y muy potente, que algunos de los jueces calificaron como “con demasiado sabor a madera”.

Estas características son poco habituales en la zona donde se ha realizado el estudio, y dado el valor parcial que representa la opinión del juez-consumidor, puede que no sean atributos apreciados en un vino de acuerdo a la visión particular de los jueces consultados. Lo cual explica cómo a pesar de ser muy valorado en ciertas regiones y la fama que guardan algunos de estos vinos, ha sido tan pobremente ordenado.

Dado que los juicios son emitidos por jueces inexpertos, están ligados a valores y experiencias subjetivas. Sería recomendable acompañar al ensayo de un estudio objetivo de los componentes químicos con el fin de poder identificar los distintos compuestos que hayan podido influir en las evaluaciones finales (aromas de fruta y roble o ausencia de notas de vegetal verde, compuestos sulfurosos etc).

5. Conclusión

A lo largo de este trabajo se ha podido comprobar, con un porcentaje elevado de coincidencia entre las técnicas de análisis sensorial e inteligencia colectiva, es decir, la aptitud de la inteligencia colectiva aplicada al análisis sensorial.

Si bien es cierto que los rankings obtenidos durante los análisis estadísticos y de inteligencia colectiva diferían ligeramente entre cada prueba, estos resultados coinciden en el ordenamiento de las mejores y peores muestras.

A pesar de que este ensayo es considerado una primera aproximación a la materia, los resultados son bastante positivos. Se puede afirmar que, en conjunto y por norma general, todas las pruebas han coincidido al identificar las muestras mejor y peor ordenadas, tanto en galletas como en vinos. Además, el uso de grandes grupos de trabajo en los que existe una gran libertad en la toma de decisiones y respuestas puede afectar a los resultados de la inteligencia colectiva del grupo.

En conclusión, sería necesario continuar las investigaciones en esta materia ya que si bien presenta grandes expectativas a la hora de cumplir su función podrían plantearse una serie de cambios. Por ejemplo, podría ser interesante la elaboración de grupos jerarquizados, de menor tamaño, pero más representativos, en los que haya un mayor control sobre otro tipo de variables (como pueden ser la edad, el origen o el sexo para los jueces) y apoyados en estudios de los componentes que puedan relacionar los atributos más representativos con la ordenación final obtenida por parte de ambas series de muestras.

6. Agradecimientos

Especial agradecimiento para todo el mundo que ha colaborado en la realización del presente ensayo.

7. Bibliografía

Anzaldúa-Morales, A., (1994). La evaluación sensorial de los alimentos, 1º Ed. Acribia. Zaragoza.

Aggarwal, I., Woolley, A.W. (2014). "The effects of cognitive diversity on collective intelligence and team learning". Symposium presented at the *50th Meeting of the Society of Experimental Social Psychology*, Columbus, OH.

Alwin D.F., Krosnick J.A. (1985). The measurement of values in surveys: A comparison of ratings and rankings, *Public Opinion Quarterly*, 49, pp. 535-552

Carpenter, R.P., Lyon, D.H., Hasdell, T.A., (2009) Análisis sensorial en el desarrollo y control de la calidad de alimentos, 2º Ed. Acribia, Zaragoza.

Casini, L., Corsi, A., Goodman, S., (2009). Consumer preferences of wine in Italy applying best-worst scaling, *International Journal Wine Bus Research*, 21, 64-78

Cauvain, S. y Young, L. (2006). Baked product: science, technology and practice, *Ed Blackwell*, 1º edición.

CCOO Federación Agroalimentaria, Observatorio industrial del sector agroalimentario de Castilla y León. (2009). Estudio del sector vitivinícola en Castilla y León. Recuperado de http://www.jcyl.es/web/jcyl/AgriculturaGanaderia/es/Plantilla100Detalle/1284250174899/_/1284254367689/Redaccion. Consultado (05/06/2018)

Charters S., Pettigrew S. (2007). The dimensions of wine quality, *Food Quality and Preference*, 18, 997-1007

Chira, K., Pacella, N., Jourdes, M., Teissedre, P.L., (2011). Chemical and sensory evaluation of Bordeaux wines (Cabernet-Sauvignon and Merlot) and correlation with wine age. *Food Chemistry*. Volume 126, Issue 4, 15 June 2011, Pages 1971-1977

Colleen, N. (2007). How to substitute applesauce for oil or butter? *The nest October* (6)30.

Contract testing (2014). *Are there real differences between brands? Chocolate chip cookies*. Recuperado de <https://www.contracttesting.com/>, Consultado 25/06/2018

Denominación de Origen Calificada Rioja. Variedades de uva. Recuperado de <https://es.riojawine.com/es/8-variedades-de-vid.html>, Consultado el 05/06/2018

Doolittle, L. 2007. The effects of replacing butter with pureed eggplant on the quality of chocolate chip cookies. *F&N*. 453:11.

Drake, M.A., (2007) Sensory analysis of diary foods, *Journal of Dairy Science*, vol 90, nº 11, p 4925-4937.

Engelbart, D. (1962). Augmenting Human Intellect: A Conceptual Framework, Sección en *Team cooperation*. Prepared for Air Force Office of scientific research, Washington, Recuperado de https://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/EngelbartPapers/B5_F18_ConceptFrameworkInd.html (consultado el 23/05/2018)

Fagin, R., Kumar, R. D., Sivakumar, D. (2003) .Efficient similarity search and classification via rank aggregation. IBM Almaden Research Center. CA, USA. pp 301-312.

Falahee, M., MacRae, A.W. (1997). Perceptual variation among drinking waters: The reliability of sorting and ranking data for multidimensional scaling, *Food Quality and Preference*, 8 , pp. 389-394.

Friedman, M. (1937). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance". *Journal of the American Statistical Association*. *American Statistical Association*. 32 (200): 675–701.

Girard, B., Yuksel, D., Cliff, M.A., Delaquis, P., Reynolds, A.G., (2001). Vinification effects on the sensory, colour and GC profiles of Pinot noir wines from British Columbia, *Food Research International*, vol. 34, pp. 483-499

Goodman S. (2009). An international comparison of retail consumer wine choice, *International Journal Wine Bus Research*., 21 , 41-49

Green, B. (2015). Testing and Quantifying Collective Intelligence. Publicado por Harvard University Recuperado de <https://scholar.harvard.edu/files/bggreen/files/ci2015-paper.pdf> . (Consultado el 30/05/2018).

Hopfer H., Heymann H. (2014), Judging wine quality: Do we need experts, consumers or trained panelists? *Food Quality and Preference*, Volume 32, Part C, 221-233

Jones, G. (2003). Fat and Fat Substitutes. *Historical Materials from University of Nebraska-Lincoln*, Extension: G03-1487-A:C-1482p.

Kamphaus, R.W.; Winsor, A.P.; Rowe, E.W. y Kim, S. (2005). A history of intelligence test interpretation. In D.P. Flanagan and P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd Ed.). New York, NY: Guilford. pp. 23–38.

Klein, M y Bicharra C., (2015) High-speed idea filtering with the bag of lemons, *Decision Support Systems* 78 (2015) 39–50
Kruskal, W.H., Wallis, W.A. (1952). Use of ranks in one criterion variance analysis. *Journal of the American Statistical Association*, 47,583-621.

Kruskal, W.H., y Wallis, W.A. (1952). Use of ranks in one criterion variance analysis. *Journal of the American Statistical Association*, 47,583-621.

Landemore, H. (2012). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: *Princeton University Press*.

Lawless, H. y Heymann, H. (2010). *Sensory evaluation of food. Principles and practice*. Ed *Springer Science y Business Media*, 596 pp. New York, USA.

Lehmann E.L., (1975) *Non parametrics: statistical methods based on ranks*. Editor *McGraw-Hill*, San Francisco, USA.

Lévy, P. (1994). *Collective Intelligence: Mankind's Emerging World in Cyberspace*, Ed. *Perseus*, p. 13.

Lévy, P. (2004). *L'Intelligence collective. Pour une anthropologie du cyberspace*, Ed: La Découverte. 1ºEd., ISBN :2707126934

Lévy, P. (2015). *Inteligencia Colectiva para educadores*. En Conferencia de la Organización de Estados Iberoamericanos. Recuperado de <https://www.oei.es/> Consultado (05/05/2018)

Maio, G.R., Bell, D.W., Esses, V.M. (1996). Ambivalence and persuasion: The processing of messages about immigrant groups, *Journal of Experimental Social Psychology*, 32, pp. 513-536

Munson, J.M., McIntyre, S.H. (1979). Developing practical procedures for the measurement of personal values in cross-cultural marketing, *Journal of Marketing Research*, 16 , pp. 48-52

Nunes, F., Madureira, T., Oliveira, V., Madureira, H. (2016). The consumer trail: Applying best-worst scaling to classical wine attributes, *Wine Economics and Policy*, Volume 5, Issue 2, 78-86

Pelsmaeker, S., Schouteten J., Lagasta, S., Dewettinck, K., Gellynck, X.(2017). Is taste the key driver for consumer preference? A conjoint analysis study, *Food Quality and Preference* Volume 62, December 2017, 323-331

Ramírez-Navas, J.S., (2012). *Análisis sensorial: pruebas orientadas al consumidor*. Ed. *ReCiTe/A*. Cali. Colombia

Rodriguez, N., Guillet, M., Fortina, J., Martin, J.F.(2000). Comparing information obtained from ranking and descriptive tests of four sweet corn products, *Food Quality and Preference*, Volume 11, Issues 1–2, Pages 47-54

Sáenz-Navajas, M.P., Avizcuri, J.M., Ballester, J., Fernández-Zurbano, P., Ferreira, V., Peyron D. (2015). Sensory-active compounds influencing wine experts' and consumers' perception of red wine intrinsic quality, *LWT-Food Science and Technology*, 60, pp. 400-411

Sáenz-Navajas, M.P., Avizcuri, J.M., Echávarri, J.F., Ferreira, V., Fernández-Zurbano, P., Valentin, D.,(2016). Understanding quality judgements of red wines by experts:

Effect of evaluation condition *Food Quality and Preference*, Volume 48, Part A, Pages 216-227

Sayadi, S., Erraach, Y., Parra-López, C., Jiménez-López, J. (2013). “Valoración y percepción de la calidad del Aceite de Oliva Virgen Extra: Influencia de la formación de los consumidores”. Presentado en simposio Expoliva 2013, *Institute of Agricultural Research and Training (IFAPA)*

Schut, M.C. (2010). On model design for simulation of collective intelligence. *Information Sciences* 180 (2010) 132–155.

Stone, H. y Sidel J.L. (1993). *Sensory Evaluation Practices*, Academic Press, 338 págs

Surowiecki, J. (2005). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*, Ed. Anchor, 296 pags.

Torricella Morales, R.G, Zamora, E. y Pulido, H., (2007) *Evaluación sensorial aplicada a la investigación, desarrollo y control de la calidad en la industria alimentaria*. 2º Ed. La Habana, Cuba: *Editorial Universitaria*, 131p

UNE 87022:1992. Análisis sensorial. Utensilios. Copa para la degustación de vino.

UNE-ISO 6658:2008. Análisis sensorial de alimentos. Metodología. Guía general.

UNE-EN ISO 8589:2010/A1:2014. Análisis sensorial. Guía general para el diseño de una sala de cata.

UNE-En ISO 8587:2010 Análisis sensorial. Metodología. Ordenación.

Villanueva, N.D.M., Petenate A.J., Da Silva M.A.A.P. (2005). Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales, *Food Quality and Preference*, 16 , pp. 691-703

Watts, B.M.; Ylimarki, G.L. Jeffery, L.E. y Elias, L.G. (1989) Basic sensory methods for food evaluation. Ottawa, Canada *International development research centre.*, 170p

Woolley, A., Chabris, F., Pentland, A., Hashmi, N., Malone, T., (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups, *Science* Vol. 330, Issue 6004, pp. 686-688.

Woolley, A. y Malone, T. (2011). Defend your research: What makes a team smarter? More women. *Harvard Business Review*. 89 (6): 32–33.