



Comparación y evaluación de las pruebas de diferencia Dúo – trío, triangular, ABX e igual diferente

R. Olivas - Gastélum*

Departamento de Ingeniería Química y Alimentos, Universidad de las Américas – Puebla. San Andrés Cholula, Pue., México.

Resumen

La evaluación sensorial de alimentos es de suma importancia en la investigación y el desarrollo de alimentos. Las pruebas sensoriales existentes se deben utilizar dependiendo del tipo de información que se desee obtener. Las pruebas de diferencia se utilizan cuando se desea conocer si dos alimentos son perceptiblemente diferentes, y son ampliamente usadas tanto en la academia como en la industria por sus aplicaciones en el control de calidad, el estudio del impacto por cambios en la formulación o el proceso, la habilidad de los consumidores para discriminar entre dos productos similares, entre otras aplicaciones. Primero, se lleva a cabo una revisión de las pruebas discriminativas, y posteriormente se describen teorías de análisis de resultados que tienen como finalidad indicar si hay o no diferencias a partir de un análisis estadístico. Finalmente, se revisan los fundamentos de la modelación Thurstoniana, incluyendo el cálculo del parámetro d' . El presente artículo provee un panorama de algunas pruebas de diferencia, considerando el proceso cognitivo que se lleva a cabo en cada una de las pruebas. La consideración de todas estas variables permitirá la selección del protocolo más adecuada en investigaciones de evaluación sensorial con este tipo de pruebas.

Palabras clave: *Pruebas de diferencia, Igual-diferente, Modelación Thurstoniana, d' , Distribución binomial.*

Abstract

Sensory evaluation has a major contribution to research and development of foods. The existent methods must be used accordingly with the desired information. Difference tests are to be used when a researcher wants to know if two confusable stimuli are perceivably different. Difference tests have important applications in quality control, reformulation studies, and consumer research, and are widely used both in industry and in education. At first, a review on difference tests, and the theories to analyze the results on difference testing are addressed. Later, Thurstonian Modelling fundamentals and d' calculations are also explained. The present work provides a revision on some difference tests, considering the cognitive strategy being held in each of the tests. The review of all the variables involved in difference testing will allow the selection of the appropriate test in sensory evaluation research.

Keywords: *Difference tests, Same-different, Thurstonian Modelling, d' , Binomial distribution.*

* Tel.: +52 222 229 2126, fax: +52 222 229 2727

Dirección electrónica: ricardo.olivasgm@udlap.mx

Introducción

La evaluación sensorial es el análisis de alimentos u otros materiales por medio de los sentidos (Anzaldúa-Morales, 1994). La palabra sensorial se deriva de latín *sensus*, que quiere decir *sentido*. La evaluación sensorial de los alimentos es una técnica de medición tan importante como los métodos químicos, físicos, microbiológicos, etc.

Las técnicas de evaluación sensorial tienen fundamento científico al igual que otros tipos de análisis al ser respaldadas por diversas áreas de estudio como la estadística y la psicología. El estudio sensorial es de suma importancia y puede resultar en una poderosa herramienta, si tan solo se estudia seria y cuidadosamente.

La evaluación sensorial de alimentos se lleva a cabo por medio de diferentes pruebas, dependiendo del tipo de información que se busque obtener. Existen tres tipos principales de pruebas: las pruebas afectivas, las de discriminación, y las descriptivas. Las pruebas afectivas son aquellas que buscan establecer el grado de aceptación de un producto a partir de la reacción del juez evaluador. Por otro lado, las pruebas de discriminación son aquellas en las que se desea establecer si dos muestras son lo suficientemente diferentes para ser catalogadas como diferentes. Finalmente, las pruebas descriptivas intentan definir las propiedades de un alimento y medirlas de la manera más objetiva posible (Anzaldúa-Morales, 1994). Cada tipo de prueba busca obtener información de una o varias muestras de alimento, no obstante, el tipo de información a obtener es muy diferente para cada una de ellas.

Para lo que ha sido denominado Evaluación Sensorial I, los sentidos humanos se utilizan como una herramienta analítica; por ejemplo, la nariz se utiliza como una

alternativa a un cromatógrafo de gases para detectar compuestos químicos volátiles. El objetivo de este tipo de evaluación es estudiar las propiedades de los alimentos, no la capacidad de los consumidores de detectar diferencias. En este caso es lógico que se busque utilizar la prueba más poderosa sensible y estadísticamente, al igual que en el laboratorio se busca utilizar el equipo analítico más sensible (O'Mahony y Rousseau, 2002).

En la Evaluación Sensorial II, se busca predecir que tan buenos son los consumidores para discriminar en condiciones naturales de consumo o uso. El objetivo de este tipo de evaluación es estudiar la sensibilidad de los consumidores para diferenciar entre dos productos. La sensibilidad de la prueba a utilizar se debe nivelar con las condiciones normales de consumo, por lo que no necesariamente se debe utilizar las pruebas más sensibles (O'Mahony y Rousseau, 2002). En el desarrollo de esta investigación se hace una revisión detallada de algunas pruebas discriminativas.

Revisión bibliográfica

Pruebas de diferencia

El área de las pruebas de discriminación se ha vuelto recientemente un área activa de investigación en evaluación sensorial (Rousseau *et al.*, 2002; Ishii *et al.*, 2007). Las pruebas discriminativas se deben usar cuando un investigador desea determinar si dos muestras son perceptiblemente diferentes (Peryam, 1958; Amerine *et al.*, 1965; Stone y Sidel, 1993, O'Mahony y Rousseau, 2002). Es posible que dos muestras tengan formulaciones químicamente diferentes pero los humanos sean incapaces de percibir la diferencia. El desarrollo de productos se basa

en esta posibilidad, al reformular los ingredientes de los productos tratando de que el consumidor no detecte diferencia alguna. De la misma manera, cuando se busca reformular para crear un producto nuevo o mejorado, es deseable que el consumidor detecte diferencia entre el producto nuevo y el ya existente (Lawless y Heymann, 1999).

Las pruebas discriminativas son concebidas como pruebas simples, sin embargo, la teoría asociada a éstas las hace más complejas de lo que comúnmente se cree. Estas pruebas son copiosamente aplicadas a alimentos y son ampliamente usadas tanto en la academia como en la industria por sus diversas aplicaciones; son de utilidad en el control de calidad, en el estudio del impacto por cambios en la formulación o el proceso, en determinar la habilidad de los consumidores para discriminar entre dos productos similares, entre otras aplicaciones (Lee *et al.*, 2007).

Las pruebas de discriminación son de mayor utilidad cuando se tiene solamente dos productos. Esto no debe confundirse con el número de muestras que utilizan los distintos procedimientos, ya que hay pruebas que presentan más de dos muestras a los evaluadores pero evalúan diferencia entre dos productos (e.g. dúo-trío, triangular). Es posible realizar pruebas de diferenciación de más de dos productos, pero no son eficientes y carecen de fondo estadístico (Lawless y Heymann, 1999).

Existen dos tipos de pruebas discriminativas; las primeras buscan establecer si hay o no diferencia entre dos muestras, independientemente de la razón o atributo por la cual se podría generar ésta. De la misma manera, también existen pruebas de diferencia que identifican un atributo o característica como la fuente de posible diferencia (e.g. dulzura, amargor, sabor a cocido). Si el experimentador conoce la

fuentes de la diferencia entre las dos muestras puede utilizar cualquiera de los dos tipos. Sin embargo, si la diferencia es debida a más de un atributo, o si el origen de la diferencia no es claro, es necesario utilizar la segunda manera (O'Mahony y Rousseau, 2002).

Lo anterior es observable en estudios donde se pretende reformular algún producto (e.g. cambio de proveedor de harina en la producción de pasteles). La variación que la reformulación produce puede extenderse a más de un atributo, por lo que evaluar el cambio en un solo atributo puede menospreciar el efecto real que la reformulación tiene. De igual manera, cuando se desea hacer estudios sobre consumidores, es posible que la inhabilidad de éstos para identificar y describir atributos específicos o la posible noción errónea de lo que un atributo significa equivoque parcial o totalmente los resultados (O'Mahony y Rousseau, 2002).

Las pruebas de diferencia que indican atributos específicos a evaluar, como las de selección forzada con 2 y 3 alternativas (conocidas en la literatura como 2-AFC y 3-AFC respectivamente, por sus siglas en inglés), son más poderosas que aquellas que no los especifican como la dúo – trío y la triangular (O'Mahony y Rousseau, 2002). Pese a esto, cuando se trabaja con jueces no entrenados en detectar atributos específicos (e.g. consumidores), es comprensible e incluso justificable, utilizar pruebas que no identifican atributos específicos.

Un aspecto importante en la selección de la prueba a utilizar es que la estrategia cognitiva para los dos tipos de pruebas es diferente, por lo tanto, éstas tenderán a ser mejores o peores en encontrar diferencias dependiendo de la prueba específica que se esté utilizando, aún cuando la diferencia real entre dos muestras sea constante (O'Mahony y Rousseau, 2002).

Para entender los mecanismos cognitivos que se llevan a cabo en las pruebas de diferencia, es necesario considerar diversos factores como la desviación de respuesta, la adaptación y los criterios de decisión involucrados. Dado que el presente estudio pretende evaluar metodologías sin especificación de atributos utilizando consumidores, se hace una revisión más detallada acerca de este tipo de pruebas.

Tipos de pruebas discriminativas sin especificación de atributo como fuente de variación

Prueba igual – diferente. Esta prueba consiste en determinar si dos muestras difieren o no, sin especificar la o las dimensiones de la diferencia. Es de utilidad cuando se evalúa cambios en la formulación que pueden afectar más de un parámetro, además de ser estadísticamente más poderosa que las pruebas tradicionales dúo-trío y triangular, cuando se utiliza su versión larga (Ennis, 2001). Un ejemplo de su uso es observable en un estudio sobre dos pasteles idénticos en formulación excepto por la cantidad de azúcar. Además del dulzor del pastel, es probable que otros parámetros como la textura o el color de la corteza cambien tras la reformulación. Sería incorrecto evaluar solamente el cambio en el dulzor, ya que subestimaría la diferencia real que existe entre los dos productos.

En el desarrollo de esta prueba se le presenta al juez un par de estímulos y se le pide que simplemente indique si los percibe como iguales o diferentes. Es importante mencionar que existen dos maneras de presentar esta prueba. La versión corta involucra la presentación de un par de muestras, ya sean iguales o diferentes. La versión larga comprende la presentación de dos pares de muestras, uno de muestras iguales y otro de muestras diferentes; estos dos pares se le presentan al juez sucesivamente y sin que éste sepa que uno de los pares es de muestras iguales y otro de diferentes (Lee *et al.*, 2007; Kim *et al.*, 2006). Un ejemplo del cuestionario de esta prueba se observa en la figura 1.

Esta prueba tiene ocho posibles secuencias de presentación (AA AB, AA BA, AB AA, BA AA, BB AB, BB BA, AB BB y BA BB) las cuales deben ser presentadas en igual número y de manera aleatoria entre los jueces (Kim *et al.*, 2006).

Prueba Triangular. En esta prueba se presentan tres muestras simultáneamente, dos de ellas son idénticas y una es de una formulación diferente. El panelista debe indicar cual de las tres es la muestra diferente. En algunas versiones de la prueba, el juez debe indicar las dos muestras que son iguales (Helm y Trolle, 1946). A pesar de que no existen estudios acerca de si es mejor

Fecha _____			
Ante usted hay dos pares de muestras.			
Pruebe las muestras del primer par en la secuencia que es presentada, de izquierda a derecha. Indique si las muestras son iguales o diferentes.			
Proceda de la misma manera con el segundo par.			
Par			
1	_____	Iguales	Diferentes
2	_____	Iguales	Diferentes

Fig. 1. Cuestionario para la prueba Igual – Diferente (adaptado de O'Mahony y Rousseau, 2002).

una versión u otra, y que de manera lógica las dos versiones son iguales, puede existir una diferencia psicológica involucrada (O'Mahony, 1995). Al igual que con la prueba igual-diferente, esta prueba permite al investigador conocer si existe diferencia perceptible entre dos productos sin tener que especificar la naturaleza de la posible diferencia (Lawless y Heymann, 1999).

La hipótesis nula para la prueba triangular establece que la probabilidad de escoger la muestra diferente cuando no existe diferencia entre las muestras es de uno en tres ($H_0: P_t = 1/3$). Para esta prueba, existen seis posibles secuencias de presentación de las muestras (AAB, ABA, BAA, BBA, BAB, ABB) que deben ser presentadas a los jueces en igual número y de manera aleatoria (Kim *et al.*, 2006). Un ejemplo del cuestionario de esta prueba se observa en la figura 2.

Para esta prueba, hay dos formatos a seguir: con referencia constante o con referencia balanceada. Cabe recalcar que para el juez evaluador no existe diferencia entre ambos formatos.

El formato de referencia constante, todos los panelistas reciben la misma muestra referencia, dando como consecuencia dos secuencias de presentación ($R_A AB$, $R_A BA$). Por otro lado, cuando se usa referencia balanceada, la mitad de los panelistas reciben una muestra como referencia y la otra mitad reciben la otra, obteniendo así cuatro secuencias de presentación ($R_A AB$, $R_A BA$, $R_B AB$, $R_B BA$). Este último método es de utilidad cuando ambos productos son prototipos y los evaluadores no están familiarizados con ninguno de estos o cuando la cantidad del producto más conocido no es suficiente para hacer la prueba con referencia

Fecha _____

Ante usted hay tres muestras. Dos de ellas son iguales entre si.
Pruebe las muestras e indique con un círculo cuál es la muestra diferente.

Fig. 2. Cuestionario para la prueba Triangular (adaptado de O'Mahony y Rousseau, 2002).

Prueba Dúo-Trío. En la prueba Dúo – Trío se presentan tres muestras simultáneamente al juez. Una de éstas está identificada como referencia y es idéntica a una de las dos muestras identificadas con código. La tarea del juez es identificar la muestra codificada idéntica a la referencia. Al igual que la prueba triangular, permite identificar si hay diferencia entre dos productos, pero no indica en qué atributo difieren.

constante (Lawless y Heymann, 1999). Un ejemplo del cuestionario de esta prueba se observa en la figura 3.

Prueba ABX. La prueba ABX es un ejercicio de emparejamiento a la muestra. El panelista recibe dos muestras de referencia, una siendo el control y otra la muestra “modificada” que generalmente tiene un cambio en su formulación o en su procesamiento. Además, recibe una muestra

Fecha _____

Frente a usted hay una muestra de referencia, marcada con R, y dos muestras marcadas con claves. Una de las muestras es idéntica a R y la otra es diferente. Pruebe primero la muestra de referencia, y después las otras muestras en el orden en que son presentadas, de izquierda a derecha. Indique con un círculo el número de la muestra más parecida a la muestra de referencia.

Referencia _____

Fig. 3. Cuestionario para la prueba Dúo – Trío (adaptado de O'Mahony y Rousseau, 2002).


X que es igual a una de las referencias expuestas. El juez deberá indicar a qué muestra de referencia es idéntica (Huang y Lawless, 1998; Lawless y Heymann, 1999; MacMillan y Creelman, 1991). En esencia, éste ejercicio se asemeja a una prueba dúo – trío en reversa. En teoría, al recibir dos muestras referencia, el juez inspecciona las dos referencias y descubre la naturaleza de la diferencia entre ambas, si es que la hay (Huang y Lawless, 1998). Al ser presentadas todas las diferencias al juez, la prueba debería tener las mismas ventajas que las pruebas duales tradicionales (O'Mahony *et al.*, 1986). Un ejemplo del cuestionario de esta prueba se observa en la figura 4.


prueba esté aventajada por el hecho de que solo se evalúa una muestra desconocida, induciendo a menos fatiga sensorial o adaptación. La naturaleza de la diferencia no es especificada a los panelistas, lo que representa un desafío a éstos para descubrirla, no obstante, la variación natural característica de los alimentos podría representar una falsa señal y atraer la atención de los jueces sobre características que no son realmente diferentes entre las muestras (Ennis y Mullen, 1986b).

Numerosos estudios se llevan a cabo en la actualidad para definir los métodos de discriminación más poderosos, con el fin de

Fecha _____

Frente a usted hay dos muestras de referencia marcadas con A y B respectivamente, y una muestra X. La muestra X es idéntica a A o a B. Indique con un círculo a qué referencia es idéntica la muestra X.

A


B



X


Fig. 4. Cuestionario para la prueba ABX (adaptado de O'Mahony y Rousseau, 2002).

El periodo de inspección de las muestras referencia puede servir como periodo de “calentamiento”. Asimismo, es posible que la

disminuir la imprecisión de resultados o conclusiones. Entre más poderosa es una prueba, mayor la probabilidad de que ésta

encuentre diferencia entre dos muestras cuando la haya (Rousseau *et al.*, 2002).

Análisis de resultados de las pruebas discriminativas

Existen varios métodos tradicionales para analizar los datos obtenidos de pruebas discriminativas. Éstos asumen que el juez fue obligado a escoger una respuesta, es decir, escogieron una respuesta aún cuando no supieran con precisión la respuesta. Esta suposición es de suma importancia para que el análisis estadístico de los resultados tenga validez.

Distribución binomial. Este análisis permite al investigador determinar si el resultado del estudio es debido al azar o si los panelistas realmente percibieron diferencias entre las muestras. La ecuación 1 calcula la probabilidad de acierto (decisión correcta, p), o la probabilidad de fracaso (decisión incorrecta, q) (Lawless y Heymann, 1999):

$$P(y) = \frac{n!}{y!(n-y)!} p^y p^{n-y} \quad \text{Ec. 1}$$

Donde n es el número total de juicios, y es el número de aciertos y p es la probabilidad de acertar por azar.

A partir de esta ecuación, Roessler *et al.* (1978) publicaron tablas para cada prueba en las que a partir del número total de jueces se indica el número de juicios correctos mínimo para indicar diferencia significativa entre dos productos.

Prueba Chi-cuadrada (χ^2) ajustada. Este método permite comparar frecuencias observadas contra frecuencias esperadas hipotéticamente. En la ecuación de cálculo es necesario corregir la continuidad ya que esta distribución es continua mas las frecuencias observadas son números enteros, es decir, no

es posible que media persona consiga la respuesta correcta. El estadístico chi-cuadrada se estima a partir de la ecuación 2:

$$\chi^2 = \left[\frac{(|O_1 - E_1|)^2 - 0.5}{E_1} \right] + \left[\frac{(|O_2 - E_2|)^2 - 0.5}{E_2} \right] \quad \text{Ec. 2}$$

Donde O_1 corresponde al número de respuestas correctas observadas, O_2 corresponde al número de respuestas incorrectas observadas, E_1 es el número de correctas esperado, igual al número de evaluaciones multiplicado por la probabilidad de tener una respuesta correcta ($p = 1/2$ para dúo – trío, $p = 1/3$ para triangular, etc.) y E_2 es el número de incorrectas esperado, igual al número de evaluaciones multiplicado por la probabilidad de obtener una respuesta incorrecta por azar ($q = 1/2$ para dúo trío, $q = 2/3$ para triangular, etc.).

Con esta información y utilizando una tabla de χ^2 (Desarrolladas por Fisher y Yates en O'Mahony, 1986), es posible analizar los resultados. Para utilizar correctamente las tablas, es importante considerar que a partir de que se evalúan dos productos, los grados de libertad de la prueba son igual a uno ($gl = \text{número de elementos} - 1$).

Distribución normal y prueba Z. Es posible utilizar el área bajo la curva de la probabilidad normal para estimar la probabilidad de que azar en este tipo de pruebas. Las tablas asociadas con la curva normal utilizan áreas específicas asociadas con valores específicos de la desviación normal (z). Stone y Sidel (1978) propusieron una ecuación para obtener el valor de z específico para pruebas de diferencia:

$$Z = \frac{X - np - 0.5}{\sqrt{npq}} \quad \text{Ec. 3}$$

En esta ecuación, X es el número de respuestas correctas, n es el número total de respuestas y p y q son la probabilidad de obtener una respuesta correcta e incorrecta respectivamente, por azar (dependiente de la prueba que se utilice: $\frac{1}{3}$ para triangular, $\frac{1}{2}$ para dúo – trío, etc.). En esta ecuación también se observa el factor de corrección a la continuidad, 0.5. Este valor es necesario ya que es imposible que la mitad de un juez esté correcta y la otra mitad incorrecta. Con el valor de z es posible analizar la probabilidad de que el juez está tomando la decisión por azar, utilizando las tablas correspondientes (O'Mahony, 1986).

Problemas en las pruebas de discriminación. Existen dos tipos de errores que se pueden cometer al probar una hipótesis nula (H_0 para cualquier prueba). El primero de estos, el Error Tipo I (α) ocurre cuando se rechaza la hipótesis nula cuando en realidad es cierta, es decir, decir que dos productos son percibidos como diferentes cuando en realidad no son perceptiblemente diferentes. El Error Tipo II (β) se refiere al riesgo de no encontrar una diferencia cuando en realidad la hay. El poder de una prueba está definido como $1 - \beta$ (Lawless y Heymann, 1999). En otras palabras, el poder de una prueba es la probabilidad de que la hipótesis nula sea rechazada cuando la hipótesis nula es falsa y el tamaño de muestra y la probabilidad de error Tipo I (α) fueron especificadas. El fracaso para lograr esto es un error Tipo II (β) (Ennis, 1993).

Un error típico en las pruebas discriminativas es no saber lo que realmente significan los resultados o la interpretación incorrecta de éstos. Si un estudio de discriminación entre dos productos es llevado a cabo correctamente y se concluye que no hay diferencia entre estos productos, es innecesario realizar un estudio de preferencia entre éstos; si la diferencia entre ambos es imperceptible, ninguna de las

muestras será realmente preferida sobre la otra. No obstante, lo anterior no funciona al revés. Cuando se lleva a cabo un estudio de preferencia y en éste no hay diferencia significativa entre ambos productos, no significa que las muestras son diferentes entre sí. El resultado del estudio indica que las dos muestras tienen el mismo nivel de agrado/desagrado mas no que son iguales entre si (Lawless y Heymann, 1999). De hecho, existen investigaciones acerca de las pruebas de preferencia que cuestionan el análisis de los datos de estas pruebas cuando existe la opción de “no preferencia” (Alfaro-Rodríguez *et al.*, 2005, Angulo y O'Mahony, 2005).

Como se mencionó anteriormente, en el análisis tradicional de resultados de las pruebas de diferencia se llevaba a cabo un análisis estadístico correspondiente al número de respuestas correctas e incorrectas y se llega a una conclusión. Sin embargo, es necesario analizar más a fondo lo que sucede realmente en este tipo de pruebas para entender la sensibilidad relativa de las pruebas y el proceso cognitivo que lleva a cabo el juez evaluador cuando responde.

El problema central en las pruebas de diferencia es la desviación de respuesta, que consiste en que cuando un juez que puede discriminar entre dos muestras, reporte que lo puede hacer (O'Mahony y Rousseau, 2002). La naturaleza de este problema tiene que ver con la prueba que se utilice.

Modelación Thurstoniana

El trabajo de Thurstone (1927) fue capaz de proveer un análisis de resultados más apropiado a pruebas de discriminación y otras pruebas en el análisis sensorial. En este análisis se calcula el valor de d' como índice para describir el grado de diferencia percibido entre dos productos. A mayor valor de d' , mayor la diferencia entre éstos.

Numerosos estudios han sido llevados a cabo basándose en este análisis para el estudio de diferentes pruebas sensoriales y su uso en la determinación de pequeñas diferencias sensoriales entre dos productos (Hautus y Irwin, 1995; Stillman y Irwin, 1995; Huang y Lawless, 1998; Masuoka *et al.*, 1995; Rousseau y O'Mahony, 1997, 2000, 2001; Rousseau *et al.*, 1998; Rousseau *et al.*, 1999; Rousseau *et al.*, 2002). La mayoría de estos estudios se han realizado bajo condiciones controladas en laboratorio.

Las ideas principales detrás de la modelación Thurstoniana son dos: por un lado, cada vez que se prueba un producto, su sabor varía en intensidad, ya sea como resultado de efectos fisiológicos o por falta de homogeneidad en la muestra. Por otro lado, en las pruebas de diferencia existe una regla de decisión o estrategia cognitiva para tomar la decisión (O'Mahony y Rousseau, 2002).

Cuando un alimento es probado repetidamente, en ocasiones se percibirá el sabor más intensamente o menos intensamente, sin embargo, existirá una intensidad promedio que ocurrirá con mayor frecuencia. Esta variación en intensidad se debe a diversas razones. En el juez existe un nivel de ruido por parte del sistema nervioso, hay adaptación especialmente por residuales de estímulos degustados anteriormente y hay variabilidad en el número de receptores que mandan una respuesta a nivel periférico, etc. En el producto puede haber falta de homogeneidad tanto en las muestras como entre éstas (O'Mahony y Rousseau, 2002).

Independientemente del origen de la variación, ésta puede ser representada por una distribución de frecuencia continua a lo largo de un eje de intensidad de sabor (fig. 5). La intensidad al momento de probar la muestra caerá en algún lugar del eje y dicha

intensidad se repetirá más comúnmente entre más cerca de la media se encuentre.

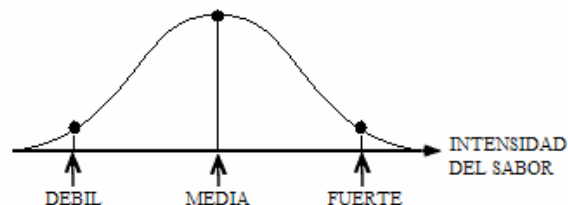


Fig. 5. Distribución de frecuencias a lo largo de un eje de intensidad de sabor representando la variación de sabor de un estímulo (O'Mahony *et al.*, 1994).

Usando este acercamiento, dos estímulos confundibles entre sí pueden ser representados por dos distribuciones que se traslapan (fig. 6). Comúnmente se asume que ambas distribuciones tienen la misma varianza, hecho que ha sido confirmado experimentalmente por diversos autores (Hautus e Irwin, 1995; O'Mahony, 1972).

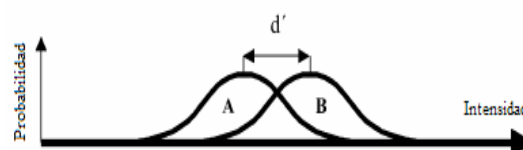


Fig. 6. Representación Thurstoniana de la diferencia/similitud entre dos estímulos (O'Mahony y Rousseau, 2002).

El grado de diferencia entre las dos muestras se denomina δ o d' (δ para poblaciones y d' para muestra experimental) y es la distancia entre las medias de las distribuciones en términos de desviaciones estándar. A mayor diferencia percibida entre dos muestras, mayor el valor de d' .

El segundo aspecto de la modelación Thurstoniana concierne a la regla de decisión. Cada prueba discriminativa tiene al menos una regla de decisión específica que el juez va a seguir para generar una respuesta. O'Mahony *et al.* (1994) indican que las dos reglas de decisión principales son

la “comparación de distancias” (en las pruebas triangular y dúo – trío) y el “desnatado” (en las pruebas 2-AFC, 3-AFC) que significa ir evaluando de mayor intensidad a menor tratando de encontrar el estímulo más significativo. Se han desarrollado tablas para relacionar la proporción de respuestas correctas con d' para diversas pruebas de discriminación (Hacker y Ratcliff, 1979; Frijters *et al.*, 1980; Ennis y Mullen, 1986a; Ennis, 1993; Ennis *et al.*, 1998; Frijters, 1982; Rousseau y Ennis, 2001)

Basándose en la teoría de Thurstone, las pruebas que utilizan la regla de decisión de “desnatado” son estadísticamente más eficientes que aquellas que realizan “comparación de distancias”, por lo tanto, éstas últimas requieren de una mayor muestra para detectar el mismo grado de diferencia.

A partir de lo descrito, parecería que las pruebas 3-AFC y 2 AFC son las más apropiadas para detectar diferencias entre dos estímulos confundibles. Esto es cierto cuando se realiza evaluación sensorial con jueces entrenados (Tipo I), ya que éstos estarán suficientemente entrenados para identificar el atributo relevante en cada prueba. No obstante, cuando no se tiene la posibilidad de identificar el atributo o se trabaja con jueces no entrenados para identificar atributos específicos (e.g. consumidores), el investigador se puede ayudar de un “calentamiento” para inducir a los jueces a identificar la diferencia. Igualmente, si la naturaleza de los estímulos no permite un proceso de “calentamiento” por exceso de fatiga, se debe utilizar pruebas sin especificación del atributo.

En la comparación de distancias, la regla de decisión más común en pruebas de diferencia sin atributo específico, existen dos estrategias cognitivas diferentes, definidas como criterios τ y β . Cuando un juez se

encuentra con varios estímulos, éste traza una línea a partir de la cual, los estímulos que perciba caerán antes o después de ésta, calificándolos como con o sin el atributo evaluado (Rousseau, 2001; Rousseau *et al.*, 1998; Lee *et al.*, 2007). Si el atributo fuera dulzura, respondería a la pregunta ¿Qué tan dulce tiene que ser el estímulo para ser llamado “dulce”? A este criterio se le conoce como criterio β , y es el utilizado en pruebas de decisión forzada con dos o tres alternativas (2-AFC y 3-AFC)

El criterio τ está orientado a la distancia que existe entre dos estímulos, y responde a la pregunta ¿Qué tan diferentes tienen que ser dos estímulos para ser considerados diferentes? (Rousseau, 2001; Rousseau *et al.*, 1998; Lee *et al.*, 2007). Las pruebas de decisión forzada como dúo-trío y triangular son procesadas por los jueces de ésta manera.

Siempre existe la posibilidad de que cuando un juez tienen ante sí los estímulos, cambie de criterio para tomar su decisión. Por ejemplo, si un juez que realiza una prueba triangular “descifra” que la fuente de la diferencia está en el dulzor, puede llevar a cabo la prueba cuestionándose “¿qué tan dulce tiene que ser una de las muestras para ser llamada dulce?”, dando como resultado global, mejores desempeños y con ello la posible detección de diferencia cuando no la hay (Lee *et al.*, 2007)

De acuerdo a Ennis (1993), algunas pruebas son más apropiadas para detectar pequeñas diferencias entre muestras. Dependiendo de la prueba sensorial aplicada, algunas reglas de decisión son más eficientes que otras, dando como resultado que un juez tenga mejores o peores desempeños con una prueba o con otra, aún cuando la diferencia entre las muestras (d') sea la misma.

Cálculo de d'

La diferencia entre dos muestras expresada en valores de d' se ha calculado para diversas pruebas, en función del número de juicios correctos e incorrectos (Ennis, 1993). No obstante, el cálculo de d' en la prueba igual-diferente es diferente ya que esta prueba puede involucrar un criterio β o τ (Hautus y Irwin, 1995, O'Mahony y Rousseau, 2002). El cálculo descrito aquí es aplicable para cuando se presenta el criterio τ . Inicialmente se calcula el tamaño del criterio τ y posteriormente se calcula d' .

idénticas (XX o YY) se consideran. Para esta situación, $d'=0$ y su desviación estándar será $\sqrt{2}$. El juez responderá que las muestras son diferentes si la diferencia percibida es mayor a τ . Por lo tanto, cualquier diferencia menor a esta, se calificará como "igual". La proporción de respuestas "igual" es equivalente al área bajo la curva entre $-\tau$ y $+\tau$ (fig. 8).

Existen tablas generadas para estudiar la relación entre el área bajo la curva de una distribución normal (por ejemplo, O'Mahony, 1986, Tabla G1), no obstante,

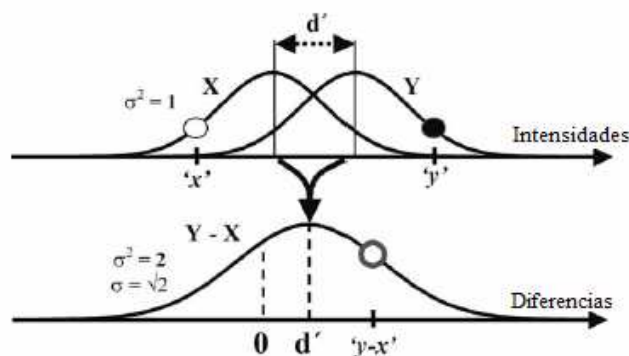


Fig. 7. Desarrollo de la distribución de diferencias en la intensidad de la percepción, ilustrado a partir de dos estímulos diferentes pero confundibles ("X" y "Y"). La percepción momentánea se ilustra con "x" y "y" (adaptado de O'Mahony y Rousseau, 2002).

De la figura 7, se observa que la diferencia entre dos distribuciones independientes de intensidades (varianza=1) se puede representar por una distribución de diferencias (varianza=2, desviación estándar= $\sqrt{2}$). Al igual que d' , τ se mide en términos de desviación estándar de las distribuciones de intensidad (valores z). Los resultados de la presentación de muestras

éstas fueron diseñadas para distribuciones con desviación estándar de 1.

Para estandarizar los datos de τ , se debe dividir todos los valores para que sean analizables por medio de las tablas normales (fig. 9).

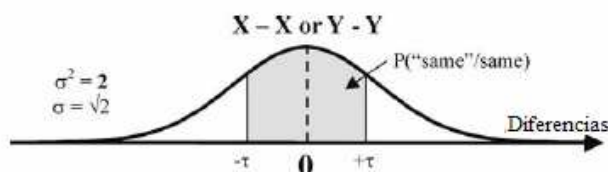


Fig. 8. Estimación del tamaño de τ a partir de presentación de muestras idénticas (adaptado de O'Mahony y Rousseau, 2002).

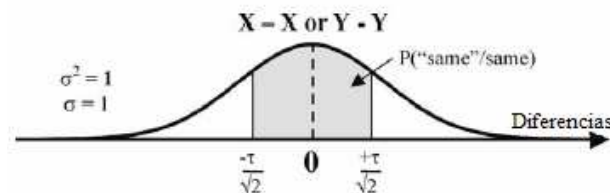


Fig. 9. Distribución estandarizada de las diferencias entre pares de muestras idénticas (adaptado de O'Mahony y Rousseau, 2002).

Una vez que se conoce el valor de τ , se calcula d' a partir de la distribución de diferencias generada para pares de muestras diferentes (XY y YX). En este caso, la media de la distribución será d' en lugar de 0 (fig. 10).

Conclusiones

Las pruebas de diferencia en la evaluación sensorial de alimentos son de suma utilidad para el análisis y desarrollo de alimentos. La pruebas de diferencia

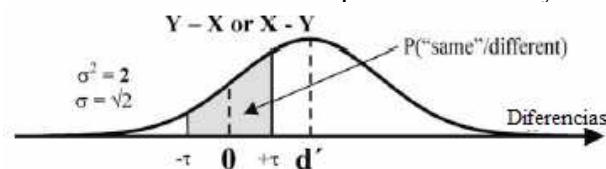


Fig. 10. Distribución de diferencias generada por muestras diferentes (adaptado de O'Mahony y Rousseau, 2002).

El cálculo de d' no es directo ya que el área bajo la curva de ésta distribución no se puede obtener directamente de tablas, no obstante, se puede obtener como la diferencia entre dos áreas: el área entre d' y $-\tau$, menos el área entre d' y $+\tau$ (fig. 11). A partir de lo anterior, se genera una ecuación de utilidad para el cálculo de d' .

aporta un fundamento estadístico que da fuerza a los resultados de las pruebas. No obstante, la selección del tipo de prueba, el tipo de jueces a utilizar y las condiciones de evaluación dictan en gran medida la efectividad y utilidad del análisis.

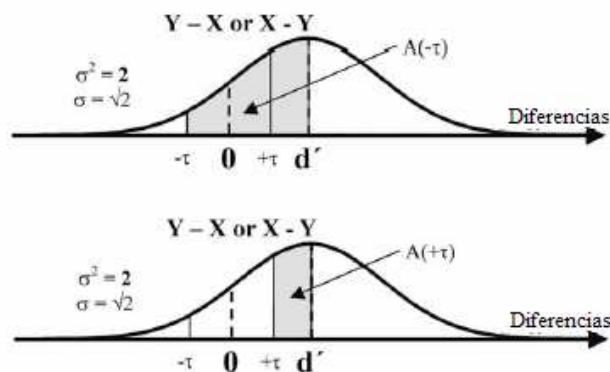


Fig. 11. Áreas utilizadas para estimar d' (adaptado de O'Mahony y Rousseau, 2002).

De acuerdo a lo revisado, cuando se busca evaluar la capacidad de consumidores de detectar diferencias entre muestras, las pruebas de diferencia sin especificación de atributo pueden ser la mejor opción, ya que se asemejan más al tipo de juicio que hacen los consumidores cuando tienen dos productos similares.

El estudio de los mecanismos cognitivos llevados a cabo al resolver un cuestionamiento de diferencia es un área de investigación en estudio que explica las diferencias entre los resultados de las diferentes pruebas de diferencia cuando se estudian dos alimentos confundibles. Se debe hacer un análisis más detallado de los procesos cognitivos en pruebas de diferencia para identificar los factores involucrados en las diferencias entre pruebas y su relación con las diferentes secuencias de presentación.

Referencias

- Alfaro-Rodriguez, H., O'Mahony, M., Angulo, O. 2005. Paired preference tests: d values from Mexican consumers with various response options. *Journal of Sensory Studies*. 20: 275 – 281.
- Amerine, M.A., Pangborn, R.M. y Roessler, E.B. 1965. Principles of Sensory Evaluation of Food. Academic. Nueva York. Citado en H.T. Lawless y H. Heymann. 1999. Sensory Evaluation of Food. Aspen Publishers, Inc. Maryland, E.E.U.U.
- Angulo, O. y O'Mahony, M. 2005. The paired preference test and the "No Preference" option: Was Odesky correct? *Food Quality and Preference*. 16: 425–434.
- Anzaldúa-Morales, A. 1994. La evaluación sensorial de los alimentos en la teoría y la práctica. Acribia. Zaragoza, España
- Ennis, D. M. 1993. The power of sensory discrimination methods. *Journal of Sensory Studies*. 8: 353–370.
- Ennis, D.M. 2001. Comunicación personal. En M. O'Mahony y B. Rousseau. 2002. Discrimination testing: a few ideas, old and new. *Food Quality and Preference*. 14: 157–164.
- Ennis, D.M. y Mullen, K. 1986a. A multivariate model for discrimination methods. *Journal of Mathematical Psychology*. 30: 206 – 219.
- Ennis, D.M. y Mullen, K. 1986b. Theoretical aspects of sensory discrimination. *Chemical Senses*, 11. Citado en H.T. Lawless y H. Heymann. 1999. *Sensory Evaluation of Food*. Aspen Publishers, Inc. Maryland, E.E.U.U.
- Ennis, J. M., Ennis, D.M., Yip, D. y O'Mahony, M. 1998. Thurstonian models for variants of the method of tetrads. *British Journal of Mathematical and Statistical Psychology*. 51: 205 – 215.
- Frijters, J. E. R. 1982. Expanded tables for conversion of a proportion of correct responses (Pc) to the measure of sensory difference (d0) for the triangular method and the 3-alternative forced choice procedure. *Journal of Food Science*. 47: 139–143.
- Frijters, J. E. R., Kooistra, A., y Vereijken, P. F. G. 1980. Tables of d' for the triangular method and the 3-AFC signal detection procedure. *Perception and Psychophysics*, 27, 176–178. Citado en M. O'Mahony y B. Rousseau. 2002. Discrimination testing: a few ideas, old and new. *Food Quality and Preference*. 14: 157–164.
- Hacker, M. J. y Ratcliff, R. 1979. A revised table of d' for M-alternative forced choice. *Perception and Psychophysics*, 26, 168 – 170. Citado en M. O'Mahony y B. Rousseau. 2002. Discrimination testing: a few ideas, old and new. *Food Quality and Preference*. 14: 157–164.

- Hautus, M. J., y Irwin, R. J. 1995. Two models for estimating the discriminability of foods and beverages. *Journal of Sensory Studies*. 10: 203–215.
- Huang, Y. T., y Lawless, H. T. 1998. Sensitivity of the ABX discrimination test. *Journal of Sensory Studies*. 13: 229–239.
- Helm, E. & Trolle, B. 1946. Selection of a taste panel. *Wallerstein Lab. Communications*, 9, 18 1-94. En M. O'Mahony. 1995. Who told you the triangle test was simple?. *Food Quality and Preference*. 6: 227–238.
- Ishii, R., Kawaguchi, H., O'Mahony M., Rousseau, B. 2007. Relating consumer and trained panel's discriminative sensitivities using vanilla flavored ice cream as a medium. *Food Quality and Preference*. 18: 89 – 96.
- Kim, H.-J., Jeon, S.J., Kim, K.-O., O'Mahony, M. 2006. Thurstonian models and variance I: experimental confirmation of cognitive strategies for difference tests and effects of perceptual variance. *Journal of Sensory Studies*. 21: 465–484.
- Lawless, H.T. y Heymann, H. 1999. *Sensory Evaluation of Food*. Aspen Publishers, Inc. Maryland, E.E.U.U.
- Lee, H. S., van Hout, D., Hautus, M., O'Mahony, M. 2007. Can the same – different test use a β – criterion as well as a τ – criterion? *Food Quality and Preference*. 18: 605 – 613.
- MacMillan, N. A., Creelman, C. D. 1991. *Detection theory: A user's guide*. Cambridge University Press, Cambridge, Reino Unido.
- Masuoka, S., Hatjopoulos, D., y O'Mahony, M. (1995). Beer bitterness detection: testing Thurstonian and Sequential Sensitivity Analysis models for triad and tetrad methods. *Journal of Sensory Studies*, 10, 295–306.
- O'Mahony, M. 1972. Salt taste sensitivity: a signal detection approach. *Perception*. 1: 459 – 464.
- O'Mahony, M. 1986. *Sensory Evaluation of Food: Statistical methods and procedures*. Marcel Dekker Inc. Nueva York, E.E.U.U.
- O'Mahony, M. 1995. Who told you the triangle test was simple? *Food Quality and Preference*. 6: 227–238.
- O'Mahony, M., Masuoka, S., y Ishii, R. 1994. A theoretical note on difference tests: models, paradoxes and cognitive strategies. *Journal of Sensory Studies*. 9: 247–272.
- O'Mahony, M., Wong, S.Y. y Odbert, N. 1986. Sensory difference tests: Some rethinking concerning the general rule that more sensitive tests use fewer stimuli. *Lebesn – Wiss. u – Technol*, 19, 93 – 94. Citado en M. O'Mahony, S. Masuoka y R. Ishii. 1994. A theoretical note on difference tests: models, paradoxes and cognitive strategies. *Journal of Sensory Studies*. 9: 247–272.
- O'Mahony, M. y Rousseau, B. 2002. Discrimination testing: a few ideas, old and new. *Food Quality and Preference*. 14: 157–164.
- Peryam, D.R. 1958. Sensory difference tests. *Journal of Food Technology*. 12: 231 – 236.
- Roessler, E.B., Pangborn, R.M., Sidel, J.L., y Stone, H. 1978. Expanded statistical tables for estimating significance in paired – preference, paired – difference, duo – trio and triangle tests. *Journal of Food Science*. 43: 940 – 941.
- Rousseau, B. 2001. The b-strategy: an alternative and powerful cognitive strategy when performing sensory discrimination tests. *Journal of Sensory Studies*, 16, 301–318.
- Rousseau, B., y Ennis, D. M. 2001. A Thurstonian model for the dual-pair (4IAX) discrimination method. *Perception and Psychophysics*. 63: 1083–1090.

- Rousseau, B., Meyer, A., y O'Mahony, M. 1998. Power and sensitivity of the same-different test: comparison with triangle and duotrio methods. *Journal of Sensory Studies*. 13: 149-173.
- Rousseau, B., y O'Mahony, M. 1997. Sensory difference tests: Thurstonian and SSA predictions for vanilla flavored yogurts. *Journal of Sensory Studies*. 12: 127-146.
- Rousseau, B., y O'Mahony, M. 2000. Investigation of the effect of within-trial retasting and comparison of the dual-pair, same-different and triangle paradigms. *Food Quality and Preference*. 11: 457-464.
- Rousseau, B., y O'Mahony, M. 2001. Investigation of the dual-pair method as a possible alternative to the triangle and same-different tests. *Journal of Sensory Studies*. 16: 161-178.
- Rousseau, B., Rogeaux, M., y O'Mahony, M. 1999. Mustard discrimination by same-different and triangle tests: Aspects of irritation, memory and t criteria. *Food Quality and Preference*. 10: 173-184.
- Rousseau, B., Stroh, S., y O'Mahony, M. 2002. Investigating more powerful discrimination tests with consumers: effects of memory and response bias. *Food Quality and Preference*. 13: 39-45.
- Stillman, J. A., y Irwin, R. J. 1995. Advantages of the same-different method over the triangular method for the measurement of taste discrimination. *Journal of Sensory Studies*. 10: 261-272.
- Stone, H. y Sidel, J. L. 1978. Computing exact probabilities in sensory discrimination tests. *Journal of Food Science*. 43: 1028 - 1029.
- Stone, H. y Sidel, J. L. 1993. *Sensory Evaluation Practices*. 2da Ed. Academic. E.E.U.U.
- Thurstone, L. L. 1927. A law of comparative judgment. *Psychological Review*, 34, 273 - 286. Citado en B. Rousseau y M. O'Mahony. 2002. Discrimination testing: a few ideas, old and new. *Food Quality and Preference*. 14: 157-164.