

# Aplicación del modelo de Thurstone a las pruebas sensoriales de diferencia

*Ofelia Angulo y Michael O'Mahony*

Unidad de Investigación y Desarrollo en Alimentos, Instituto Tecnológico de Veracruz, Veracruz, México.  
Department of Food Science and Technology University of California, Davis, California, E.E.U.U.

**RESUMEN.** En las pruebas de diferencia, el desempeño de los jueces varía en función de la prueba utilizada. Esta diferencia se explica a partir del conocimiento sobre la forma en que el cerebro procesa la información que recibe de los sentidos, particularmente el sentido del gusto. En algunas pruebas de diferencia, la información se procesa de manera más eficiente que en otras. El modelo de Thurstone describe las estrategias del cerebro para procesar la información en función del método utilizado. Esta revisión introduce las ideas y los conceptos de este modelo en las mediciones sensoriales derivadas con las pruebas de diferencia. Los conceptos de la variabilidad del estímulo y la estrategia cognitiva correspondiente a cada método de diferencia son la base para el cálculo de la magnitud de la diferencia percibida entre dos estímulos, denominada  $d'$ . Esta medición es fundamental e independiente de la metodología sensorial utilizada.

**Palabras clave:** Evaluación sensorial, pruebas de diferencia, modelo Thurstoniano, teoría de detección de señales, estrategias cognitivas.

**SUMMARY. Thurstone model application to difference sensory tests.** Part of understanding why judges perform better on some difference tests than others requires an understanding of how information coming from the mouth to the brain is processed. For some tests it is processed more efficiently than others. This is described by what has been called Thurstonian modeling. This brief review introduces the concepts and ideas involved in Thurstonian modeling as applied to sensory difference measurement. It summarizes the literature concerned with the theorizing and confirmation of Thurstonian models. It introduces the important concept of stimulus variability and the fundamental measure of sensory difference:  $d'$ . It indicates how the paradox of discriminatory non-discriminators, which had puzzled researchers for years, can be simply explained using the model. It considers how memory effects and the complex interactions in the mouth can reduce  $d'$  by increasing the variance of sensory distributions.

**Key words:** Sensory evaluation, difference tests, Thurstonian models, signal detection theory, cognitive strategies.

## INTRODUCCION

La ciencia progresa a través del desarrollo de teorías y modelos. Hasta hace poco, la ciencia sensorial de los alimentos se consideraba como una compilación de métodos sin fundamento teórico. Se asumía que sí el análisis estadístico de los datos era suficientemente sofisticado, esto compensaría por los problemas asociados con las mediciones sensoriales. Sin embargo, esta suposición no es válida. Independientemente de lo sofisticado del análisis estadístico, nada puede hacer que los defectos en las mediciones sensoriales inapropiadas desaparezcan. Recientemente, este error se ha puesto de manifiesto. Los investigadores han reconocido que las mediciones sensoriales presentan problemas por lo que el énfasis hoy en día es en la aplicación de modelos a la metodología sensorial. El modelo más avanzado, a la fecha, es el de Thurstone.

### Pruebas de diferencia

El modelo de Thurstone fue originalmente aplicado a las pruebas de diferencia. Estas últimas permiten determinar el grado de diferencia o la magnitud de la diferencia entre dos

estímulos confundibles, los cuales son tan parecidos que resulta difícil considerarlos como iguales o diferentes. Este tipo de mediciones sensoriales son importantes en estudios sobre vida útil, reformulación de producto, tipo de empaque, cambio de proceso de producción, etc., así como en programas de control de calidad y desarrollo de nuevos productos.

Las pruebas de diferencia no presentan la misma sensibilidad (1-16). Las investigaciones pueden resultar mostrando diferencias con una metodología y falta de diferencia con otra metodología sensorial. Hay varios reportes donde se compara el desempeño de los jueces utilizando varios protocolos sensoriales para evaluar diferencias. Fue precisamente esta discrepancia en los resultados que propició el desarrollo de teorías y modelos.

Como primer intento, se ha aplicado el modelo Thurstoniano (17,18) y la teoría de detección de señales (19, 20). El modelo Thurstoniano fue originalmente aplicado en las pruebas triangular, dúo-trío y comparación por pares o 2-AFC (prueba de selección forzada de dos alternativas por sus siglas en inglés, (2-Alternative Forced Choice) (21). Posteriormente, se han desarrollado modelos univariados y multivariados para otras pruebas de diferencia (22-47). Estos

modelos han permitido el cálculo del parámetro  $d'$  ( $d'$  prima), que es el índice Thurstoniano de grado de diferencia. El cálculo se realiza a partir de la proporción de respuestas correctas obtenida en la prueba de diferencia correspondiente. Los modelos desarrollados hicieron posible el desarrollo de tablas que relacionan la proporción de respuestas correctas y el parámetro  $d'$  (21, 34,48-52), y dicho modelo y sus aplicaciones se han discutido ampliamente (24, 53-56). La comprensión del modelo se facilita al entender los conceptos de la variabilidad de la percepción sensorial y el índice de discriminación, denominado parámetro  $d'$ .

### Suposiciones del modelo Thurstoniano

El modelo de Thurstone se basa en dos suposiciones: la variabilidad de la percepción sensorial y el establecimiento de la estrategia cognitiva propia de cada prueba sensorial. Para explicar el significado de la primera suposición debemos imaginar la degustación de estímulos. Al evaluar sensorialmente un estímulo 'X', todas las sensaciones emanadas de la degustación del estímulo son enviadas al cerebro a través del sistema nervioso, pero las sensaciones no son constantes. El sistema nervioso presenta fluctuaciones en el número de nervios estimulados por la sensación y por la intensidad de la sensación derivada de la degustación. Esto es el resultado de las diferentes interacciones que tienen lugar en la boca. El sabor del estímulo se diluye con los fluidos orales, que incluyen la saliva que se secreta a diferentes velocidades y la presencia de residuos de la degustación previa. El sistema gustativo se torna insensible a la presencia de estos componentes de tal manera que la intensidad de la degustación y la estimulación transmitida al cerebro varía. Además, el mismo estímulo puede ser heterogéneo, lo que aumenta la variabilidad a la estimulación nerviosa. Así, algunas degustaciones resultarán más intensas y otras menos intensas. Pero, la intensidad promedio de la degustación será la que se presente con mayor frecuencia. La frecuencia con que aparecen las diferentes intensidades de percepción del estímulo puede representarse a través de una distribución normal de frecuencias. La aplicación de la primera suposición del modelo Thurstoniano es la representación de la distribución de la intensidad de la percepción sensorial del estímulo. Al evaluar un segundo estímulo 'Y' diferente al primero, la intensidad de la percepción se espera sea diferente. El nuevo estímulo aumentará la cantidad de estimulación transmitida al cerebro de tal manera que en términos de la distribución de frecuencias puede verse como una nueva distribución con intensidades más altas. El juez puede fácilmente discriminar entre un estímulo y otro, las dos distribuciones de frecuencia son perceptiblemente diferentes.

Sin embargo, al comparar dos estímulos ligeramente diferentes, las distribuciones de frecuencia se superponen entre sí y los estímulos serían confundibles. En el nivel alto de

intensidad de la percepción, el estímulo se identificaría como 'Y', en intensidades bajas se identificaría como 'X'. En intensidades promedio, en donde las distribuciones se superponen, los estímulos se percibirían a veces como 'X', a veces como 'Y', sería difícil identificarlos correctamente. Evidentemente que entre más separadas se representen las dos distribuciones de frecuencia de percepción, más fácil será discriminarlos. Así, el nivel de separación de las distribuciones puede utilizarse como una medida de que tan bien los jueces discriminan entre los dos estímulos. La distancia entre las medias de cada una de las dos distribuciones constituye el parámetro de  $d'$ ; es un término adoptado por ingeniería de comunicaciones. Entre más grande es el valor de  $d'$ , mayor es el grado de diferencia entre los estímulos. La distancia entre las medias se mide en unidades de desviación estándar. Así, cuando el valor de  $d'$  es igual a 1.5 implica que las distribuciones de intensidad de percepción de los estímulos se encuentran separadas por una distancia de 1.5 desviaciones estándar. Prácticamente  $d'$  es una medida de la diferencia entre dos estímulos en términos de la variabilidad del estímulo. Para un ingeniero de comunicaciones es equivalente a la relación señal-ruido.

### Varianza de las distribuciones

La variabilidad de la percepción sensorial de acuerdo al modelo Thurstoniano puede representarse en una distribución normal (Figura 1). Dicho modelo provee mayor información al comparar la varianza de las distribuciones de la percepción sensorial de los productos evaluados. De acuerdo a las distribuciones representadas en la Figura 2, la unidad de medida de  $d'$  es la desviación estándar. Si la desviación estándar aumenta, el valor de  $d'$  disminuye. Así, para una desviación estándar igual a uno, el valor de  $d'$  puede ser igual a dos. Si la desviación estándar se duplica, la unidad de medición también se duplica y el valor de  $d'$  se reduce a la mitad. Cualquier variable experimental que presente desviaciones estándar o varianza elevada, reducirá el valor de  $d'$ . ¿Cuáles son estas variables que incrementan la varianza? En primer lugar, el grado de reacción al estímulo del sistema gustativo puede asumirse como la principal causa que genera la diferencia de percepción entre las dos distribuciones. Jueces con mayor variabilidad en el sistema nervioso, debido a fluctuaciones grandes en la velocidad de activación del sistema nervioso no relacionada con la presencia del estímulo bajo evaluación si no que sólo son estimuladas espontáneamente, presentará una distribución con varianza más grande.

Como resultado, el valor de  $d'$  disminuye. Otra forma de expresar lo mismo es decir que para que dos productos se perciban como diferentes, éstos deben presentar diferencias mayores a la variabilidad aleatoria en el sistema sensorial para un producto dado.

FIGURA 1

Distribuciones de frecuencia que representan estímulos confundibles, con un enfoque de modelo Thurstoniano ilustrando la prueba de comparación por pares

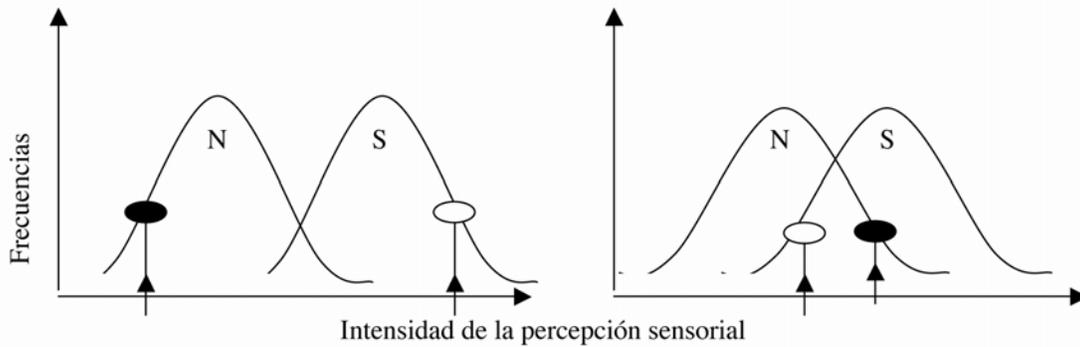
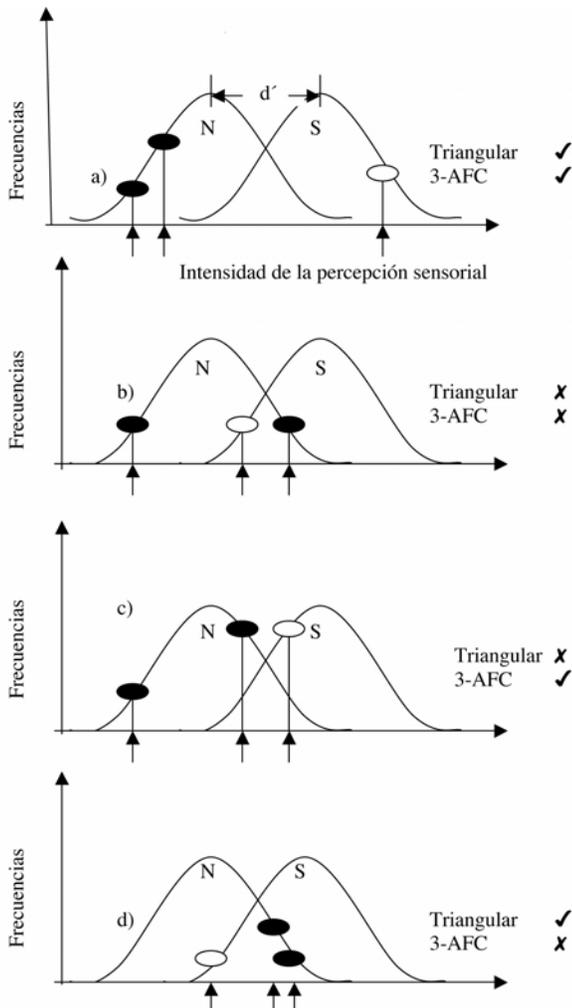


FIGURA 2

Aplicación del modelo de Thurstone a la prueba triangular y 3-AFC



En segundo lugar, las diversas interacciones que ocurren en la cavidad oral también contribuyen en la varianza de manera importante. A mayor varianza, más difícil le resulta al juez detectar diferencias entre los productos a evaluar. De igual manera, si los productos presentan variabilidad en sí mismos, debido a la mala calidad, será más difícil para el juez percibir las diferencias entre los productos. Así, las diferencias entre los productos deben ser superiores a la variabilidad de los mismos productos para reportar diferencias.

Otra fuente de variación es el resultado de la secuencia de las degustaciones de los estímulos. Por ejemplo, al degustar un alimento, seguido de otro alimento de intensidad similar o más intenso al primero, la insensibilidad provocada por el remanente de la primera degustación hará que el segundo alimento se perciba menos intenso de lo que debiera. El efecto será inverso si la secuencia de las degustaciones cambia. Sin embargo, un aumento de la salivación, en la expectoración o en la deglución disminuirá el efecto. De cualquier manera, el efecto estará más o menos presente, aumentando la variabilidad en la distribución (31, 60-63, 65-73).

El efecto de la memoria y el olvido añaden variabilidad a la distribución de percepción de los productos. Las pruebas de diferencia involucran comparar el sabor de un alimento con el recuerdo del alimento previamente degustado. Mientras más tiempo pasa entre las degustaciones, menos preciso será la variabilidad asociada a la memoria del alimento degustado previamente. Este aumento en la variabilidad producirá una reducción en el valor de  $d'$ , lo cual predice que las pruebas de diferencia con tres muestras son menos discriminatorias que las de dos estímulos (13,71,72-76).

**Aplicaciones del modelo**

No es fácil entender el concepto de  $d'$  sin antes usarlo. A fin de tener una idea de lo que representa, es útil saber que cuando  $d'$  es igual a uno representa la diferencia entre dos

estímulos que apenas se perciben como diferentes. La magnitud de la diferencia se dice que está en el umbral de percepción. El concepto se comprende mejor a través de un ejemplo en donde se aplica la prueba de diferencia de comparación por pares.

Imaginemos la situación en la que a un juez se le pide que identifique, entre dos muestras, la que percibe con mayor intensidad en dulzor. Si el juez discrimina fácilmente entre las dos muestras, todas las pruebas de comparación por pares serán correctas (100% de respuestas correctas). Si el juez no discrimina entre las muestras, su respuesta será aleatoria y obtendrá 50% de respuestas correctas. Cuando obtiene 75% de respuestas correctas se encuentra en el umbral de percepción de la diferencia; es decir entre discriminación total (100% de respuestas correctas) y respuesta aleatoria (50% de respuestas correctas). El valor de  $d'$  igual a uno corresponde al 76% de respuestas correctas en una prueba de comparación por pares. El cálculo de  $d'$  toma en cuenta las suposiciones del modelo de Thurstone (49). La Tabla 1 presenta algunos ejemplos de la relación entre el valor de  $d'$  y la proporción de respuestas correctas para diferentes pruebas de diferencia.

TABLA 1  
Valores de  $d'$  en función de la proporción de respuestas correctas y del tipo de prueba de diferencia

Proporción de respuestas correctas (%)	2-AFC	3-AFC	Dúo-Trío	Triangular
50	0	0.56	0	1.47
55	0.18	0.72	0.76	1.72
60	0.36	0.89	1.12	1.98
65	0.55	1.06	1.42	2.23
70	0.74	1.24	1.72	2.5
75	0.95	1.43	2.02	2.8
80	1.19	1.65	2.36	3.13
85	1.47	1.91	2.75	3.52
90	1.82	2.23	3.26	4.03
95	2.33	2.71	4.06	4.8
99	3.29	3.62	5.7	6.09*

\*Valor máximo en tablas corresponde a 98.71% de respuestas correctas

A través de la Figura 1 se puede presentar un ejemplo de la aplicación del modelo. Dicha figura representa las distribuciones de frecuencia de percepción de intensidades de dos estímulos confundibles, pues las distribuciones se superponen. Suponiendo que N representa el estímulo menos intenso y S representa el estímulo más intenso; cuando el juez degusta cualquier estímulo lo puede percibir en un nivel de intensidad cualquiera de la distribución correspondiente. La

tendencia más frecuente es que se perciba a un nivel promedio de intensidad en donde la frecuencia es más alta. Si nos imaginamos que la intensidad del estímulo N, representada por el óvalo negro, se percibe en un nivel bajo de intensidad y que cuando el juez degusta el estímulo S, representado por el óvalo blanco, lo percibe a un nivel de intensidad alto, al contestar la pregunta de cuál de los dos estímulos es más intenso en dulzor, indicará el estímulo S, y se le calificará correctamente (Figura 1, lado izquierdo). Al considerar nuevamente la figura 1, en esta degustación el sistema nervioso esta un poco más activo de tal manera que al degustar el estímulo N, lo percibe a un nivel más alto de intensidad del atributo (óvalo negro). Por otro lado, al degustar el estímulo S, el sistema nervioso se encuentra en un nivel de estimulación bajo, de tal manera que el nivel de intensidad de percepción se encuentra en la parte menos intensa de la distribución S (óvalo blanco). Si le pregunta sobre cuál de los dos estímulos es más intenso en dulzor, elegirá el estímulo N, dando un resultado incorrecto (Figura 1, lado derecho). Así, algunas veces el juez acierta en su respuesta, otras no. Esta es la situación durante la degustación de estímulos confundibles que normalmente son evaluados por pruebas de diferencia. Entre más se superponen las dos distribuciones (menor valor de  $d'$ ), mayor la probabilidad de error de los jueces. Al repetir este mismo ejercicio 100 mil veces a través de modelo de simulación Monte Carlo, se estimaría la proporción de veces que la prueba de comparación por pares se acierta para un valor de  $d'$  en particular. Este mismo ejercicio se repetiría para toda una serie de valores de  $d'$ . Lo que daría origen a una tabla en donde se relacionan la proporción de respuestas correctas y el valor de  $d'$  (49).

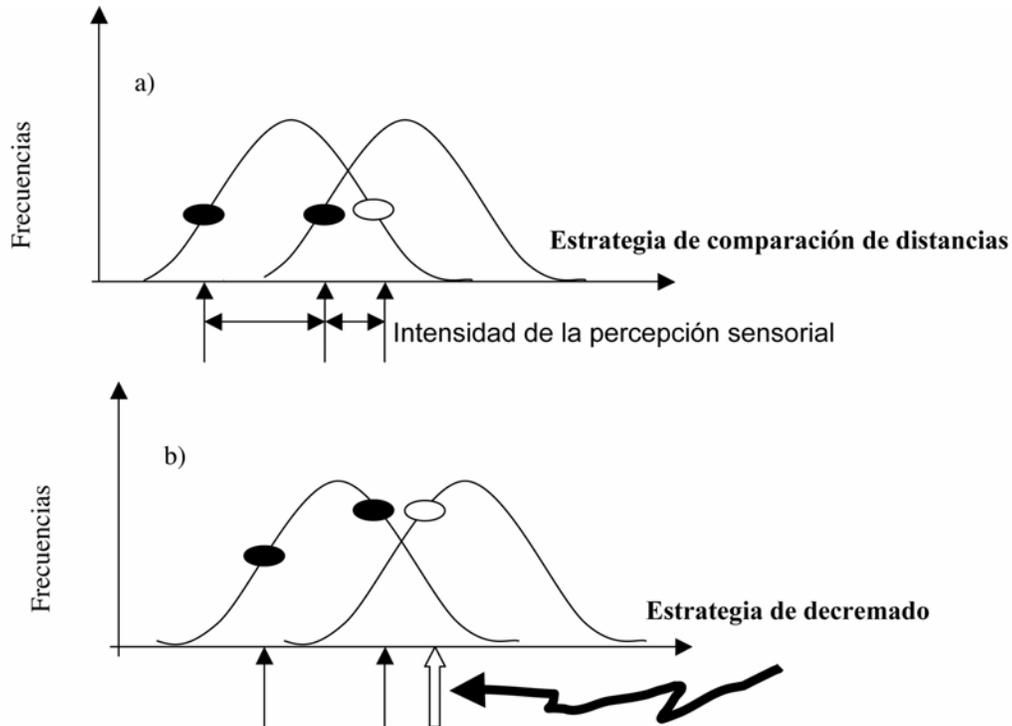
Este mismo ejercicio puede representarse para las diferentes pruebas de diferencia, dando como resultado tablas que correlacionan la proporción de respuestas correctas con el valor correspondiente de  $d'$ . Esto quiere decir que las pruebas de diferencia como la prueba de comparación por pares y la prueba triangular pueden compararse directamente aún cuando la probabilidad de respuesta al azar de ambas pruebas es diferente. El parámetro  $d'$  permite la comparación de pruebas de diferencia con probabilidades de respuesta aleatoria diferentes, algo que no se podría hacer al aplicar el modelo binomial. Por lo tanto,  $d'$  es una medición básica, fundamental pues es independiente de la metodología aplicada.

### La paradoja de las pruebas discriminatorias

Al aplicar las diferentes pruebas de diferencia para evaluar el mismo par de estímulos, algunas pruebas permiten discriminar mientras que con otras no se reportan diferencias entre los estímulos. Esta paradoja de resultados puede ejemplificarse a través de las pruebas triangular y 3-AFC (Figura 3).

FIGURA 3

Ilustración de las dos estrategias cognitivas o reglas de decisión usadas durante pruebas de diferencia: comparación de distancia vs. descremado



En la prueba triangular se presentan tres estímulos al juez; en donde dos de las muestras corresponden al mismo estímulo y una muestra corresponde al otro estímulo (muestra diferente). La tarea del juez es degustar las tres muestras e identificar la muestra que es diferente. En la prueba 3-AFC (3-alternative forced choice), la tarea es ligeramente diferente. El juez degusta las tres muestras pero ahora se le indica el atributo en el que difieren las muestras. El juez debe identificar la muestra con mayor intensidad en el atributo a evaluar. Por ejemplo, al evaluar la acidez en dos muestras de café usando la prueba 3-AFC, el juez degusta las tres muestras y se le pide identificar la muestra de café con mayor intensidad en acidez. Los resultados de la aplicación de las pruebas triangular y 3-AFC reportan mejor desempeño de los jueces cuando utilizan la prueba 3-AFC. Esta diferencia en el desempeño de los jueces se reportó por primera vez en 1937 por Abrahams y colaboradores (57) y durante muchos años no se pudo explicar. Años más tarde, se le conoció como la paradoja de las pruebas de diferencia (discriminatorias y no-discriminatorias) (8). Es posible que se atribuya el mejor desempeño de la prueba 3-AFC al hecho de que el juez conoce el atributo a evaluar. Sin embargo, ésta no es la explicación. Hay otras pruebas sensoriales en donde el juez se desempeña mejor cuando

desconoce el atributo que cambia (58). No fue sino hasta 1979 que Frijters (35, 36) encontró la solución a la paradoja, a través de los conceptos del modelo de Thurstone. La solución es la siguiente.

Si se considera nuevamente los estímulos N y S de la Figura 2, siendo dos estímulos del tipo N y un estímulo del tipo S, ligeramente más intenso. Los dos estímulos denotados con los óvalos negros pertenecen a la distribución del estímulo N, mientras que el óvalo blanco pertenece a la distribución del estímulo S. Al considerar la parte superior de la Figura 2 (Figura 2a), se observa que los dos estímulos (óvalos negros) se encuentran en la parte central y baja de la distribución N, el estímulo de mayor intensidad S, se denota por el óvalo blanco. Al aplicar la prueba triangular, el juez deberá identificar la muestra diferente y elegirá el estímulo S, pues es el más alejado de los dos estímulos representados por los óvalos negros. Al aplicar la prueba 3-AFC, deberá identificar la muestra más intensa y elegirá el estímulo S, pues es el más alejado a la derecha en el eje de intensidad de percepción. De tal manera que en este ejercicio en ambas pruebas se reportan respuestas correctas.

Al considerar las distribuciones de la Figura 2b, se observa que un estímulo se asocia con la parte izquierda de menor frecuencia de la distribución N y el otro se localiza en la parte

cercana a la frecuencia más intensa. El estímulo S por otro lado, se localiza en la parte más baja de intensidad de su distribución; en realidad se localiza a un nivel de intensidad más bajo que uno de los estímulos N. Al aplicar la prueba triangular, el juez elegirá la muestra más alejada y en este caso será el estímulo negro de la izquierda de la distribución N. Con la prueba 3-AFC, al identificar la muestra de mayor intensidad de percepción, elegirá el estímulo más alejado en el eje de intensidad, es decir el estímulo del óvalo negro. En ambos casos, la respuesta es incorrecta. Sin embargo, el error no es el mismo. Las pruebas sensoriales triangular y 3-AFC no son equivalentes.

Al considerar las distribuciones de la Figura 2c, uno de los estímulos representados con los óvalos negros se localiza en la parte baja en intensidad de la percepción del estímulo N y el segundo se localiza en la parte alta de la distribución. El estímulo S (óvalo blanco), ubica la intensidad de percepción en la parte baja de la distribución. Con la prueba triangular, el juez identificará el estímulo N ubicado en la parte baja de la distribución como la muestra diferente, pues es la muestra más alejada de las otras dos. Desde luego que esta respuesta es incorrecta. Con la prueba 3-AFC, el juez identificará la muestra S como la muestra de mayor intensidad de percepción, pues es la muestra más a la derecha en el eje de intensidad del atributo. Esta respuesta es correcta. Es así que a partir de la misma información que el cerebro recibe de los sentidos, el juez toma una decisión correcta con la prueba 3-AFC e incorrecta con la prueba triangular. Sin embargo, la sensibilidad del juez es la misma. Esta discrepancia en los resultados se debe a que el juez realiza la toma de decisión de manera diferente en cada prueba. El proceso de toma de decisión con la prueba 3-AFC es más eficiente.

Finalmente, al analizar la Figura 2d se observa que los estímulos están totalmente intercambiados. Este suceso es raro. Sin embargo, en esta situación la prueba triangular resultaría correcta y la 3-AFC incorrecta.

Al modelar estos escenarios con estímulos que presentan grado de discriminación correspondiente al valor de  $d'$  igual a uno y para una muestra de 100 000, el escenario ilustrado en el caso (a) aparece 37% de las veces; el caso (b) ocurre 32% de las veces; el caso (c) aparece 26% de las veces, y el caso (d) ocurre 5% de las veces. Así, para estímulos con el mismo grado de discriminación (igual valor de  $d'$ ), el desempeño de la prueba 3-AFC será superior al de la prueba triangular. La explicación teórica de Frijters's (35), sobre la paradoja de las pruebas de diferencia, hizo necesaria la confirmación experimental. Fue necesario demostrar experimentalmente que para el mismo grupo de jueces que degusta el mismo par de estímulos reporta proporción de respuestas correctas mayores utilizando la prueba 3-AFC que con la prueba triangular (59-63). Por ejemplo, para  $d'$  igual a uno, se espera 42% de respuestas correctas con la prueba triangular y 63% de

respuestas correctas con la prueba 3-AFC.; para  $d'$  igual a dos, la proporción de respuestas correctas será 60 y 87%, respectivamente. Estos valores están disponibles en tablas que relacionan la proporción de respuestas correctas con el valor de  $d'$  para cada prueba de diferencia (49). Es interesante recordar que la probabilidad de respuesta al azar en ambas pruebas es la misma: 33.33%. Las pruebas estadísticas que se basan en la probabilidad de respuesta al azar como la prueba binomial y  $\pm 2$  no consideran el desempeño superior de la prueba 3-AFC con respecto a la prueba triangular. El análisis basado en el modelo de Thurstone y el cálculo de  $d'$  es más revelador. La significancia estadística de  $d'$  también puede estimarse (64).

La prueba de comparación por pares y la prueba dúo-trío también pueden modelarse de la misma manera. En éstas, el desempeño de la prueba de comparación por pares es superior al de la prueba dúo-trío, aun cuando el valor de  $d'$  es el mismo. Así, para muestras con valor de  $d'$  igual a uno, se espera 58% de respuestas correctas con la prueba dúo-trío y 76% con la prueba de comparación por pares. Para  $d'$  igual a dos, la proporción de respuestas correctas será 75% para dúo-trío y 92% para comparación por pares.

### Estrategias cognitivas

¿Porqué la prueba 3-AFC y la prueba de comparación por pares permiten un desempeño superior en los jueces? La explicación se resume a través de las estrategias cognitivas que utiliza el juez en las pruebas triangular y 3-AFC (Figura 3). En la prueba triangular, el juez ubica la muestra diferente como la que más se aleja de las otras dos muestras. Esta tarea la realiza comparando las distancias entre los estímulos en el eje de intensidad de la percepción. Esta estrategia se denomina 'comparación de distancias' (Figura 3). En la prueba dúo-trío, los jueces también aplican la estrategia de comparación de distancias.

En la prueba 3-AFC, el juez elige la muestra de mayor intensidad de percepción. Esta estrategia se conoce como 'descremado' porque el estímulo de mayor intensidad es eliminado de la misma manera que se elimina la espuma de la cerveza (Figura 3). La prueba de comparación por pares utiliza la estrategia cognitiva de descremado. En el análisis de la Figura 3 se observa que la estrategia de descremado es más eficiente que la de comparación de distancias.

Cuando estos principios no se comprenden, se puede llegar a conclusiones erróneas. Si un panel reporta 90% de respuestas correctas en la prueba de comparación por pares puede concluirse que la diferencia entre los productos evaluados es más grande que la diferencia reportada por un panel con 80% de respuestas correctas usando la prueba dúo-trío. Estadísticamente hablando, esta conclusión es lógica pues ambas pruebas tienen la misma probabilidad de respuesta al azar. Sin embargo, la prueba de comparación por pares opera

con la estrategia de descremado (más eficiente) y la dúo-trío con la estrategia de comparación de distancias (menos eficiente). Para que la prueba dúo-trío resulte en un 80% de respuestas correctas, la diferencia entre las muestras, estimada por el parámetro  $d'$ , deberá ser igual a 2.36. En la prueba de comparación por pares, 90% de respuestas correctas, corresponde a una diferencia entre los estímulos de 1.82. En consecuencia, a pesar de lo esperado estadísticamente con la probabilidad de respuesta al azar, la diferencia estimada con la prueba dúo-trío es mayor que la estimada con la prueba de comparación por pares.

### Poder estadístico

La potencia estadística de una metodología se entiende como la capacidad para reportar diferencias cuando estas existen. La eficiencia de las estrategias cognitivas de las pruebas de diferencia puede ilustrarse a través de la potencia estadística. Los resultados de la medición indican que para un valor de  $d'$  igual a uno, 5% de significancia estadística y con 90% de potencia (capacidad de detectar diferencias 90% de las veces), la prueba 3-AFC requiere 21 comparaciones para declarar diferencias mientras que con la prueba triangular requiere 276 evaluaciones. El número de evaluaciones requeridas para la prueba de comparación por pares y dúo-trío es 27 y 310, respectivamente. Estos valores claramente ilustran la diferencia masiva en potencia estadística entre las pruebas sensoriales que aplican la estrategia de descremado y la estrategia de comparación de distancias.

### CONCLUSIONES

El modelo Thurstoniano permite comprender las bases de las pruebas discriminatorias o de diferencia. Uno de los aspectos más relevantes puede ser la variabilidad de la percepción de los estímulos. Es decir, un alimento que se degusta de manera rápida y repetitiva no se percibe exactamente con la misma intensidad en cada degustación. La variabilidad en la percepción sensorial constituye el principal fundamento del modelo Thurstoniano. Este supuesto provee de una medición básica de la diferencia entre dos estímulos confundibles a través del parámetro  $d'$  ( $d'$  prima), que es independiente de la prueba de diferencia utilizada. La capacidad para determinar mediciones básicas o fundamentales es esencial para cualquier disciplina científica. Las mediciones de desempeño, como la proporción de respuestas correctas no son independientes de la metodología sensorial. El desempeño del juez con la prueba 3-AFC es mejor que con la prueba triangular, mientras que la sensibilidad ( $d'$ ) para ambas pruebas es la misma. Finalmente, la consideración de la varianza de las distribuciones permite explicar las variables experimentales asociadas a la memoria y la secuencia de las degustaciones.

Los ejemplos de las distribuciones mencionados son univariados y tienen que ver con pruebas de diferencia. Pero los mismos conceptos aplican para mediciones de intensidad con escalas, para pruebas de diferencias multivariadas, para estadística multivariada o para mapeo de productos. El modelo de Thurstone es una estructura poderosa para entender los mecanismos de la medición sensorial.

### REFERENCIAS

1. Braun V, Rogeaux M, Schneid N, O'Mahony M, Rousseau B. Corroborating the 2-AFC and 2-AC Thurstonian models using both a model system and sparkling water. *Food Qual. Pref.* 15: 501-507;2004.
2. Byer AJ, Abrams D. A comparison of the triangular and two-sample taste-test methods. *Food Technol.* 7: 185-187;1953.
3. Dawson EH, Dochterman E. A comparison of sensory methods of measuring differences in food qualities. *Food Technol.* 5: 79-81;1951.
4. Filippello F. A critical comparison of the two-sample and triangular binomial designs. *Food Res.* 21: 235-241;1956.
5. François P, Sauvageot F. Comparison of the efficiency of pair, duo-trío and triangle tests. *J. Sens. Stud.* 3: 1-94;1988.
6. Gridgeman NT. Taste comparisons: Two samples or three? *Food Technol.* 9: 148-150;1955.
7. Gridgeman NT. Group size in taste sorting trials. *Food Res.* 21: 534-539;1956.
8. Gridgeman NT. A re-examination of the two-stage triangle test for the perception of sensory differences. *J. Food Sci.* 35: 87-91;1970.
9. Grim AC, Goldblith SA. Some observed discrepancies in application of the triangle test to evaluation of irradiated whole egg magma. *Food Technol.* 19: 146 ;1965.
10. Helm E, Trolle B. Selection of a taste panel. *Wallerstein Lab. Communications* 9:181-194 (1946)
11. Hopkins JW. Some observations on sensitivity and repeatability of triad taste difference tests. *Biometrics* 10: 521-530;1954.
12. Hopkins JW, Gridgeman NT. Comparative sensitivity of pair and triad flavor intensity difference tests. *Biometrics* 11: 63-68;1955.
13. Lau S, O'Mahony M, Rousseau B. Are three-sample tasks less sensitive than twosample tasks? Memory effects in the testing of taste discrimination. *Percept. Psychophys.* 66: 464-474;2004.
14. O'Mahony M, Wong SY, Odbert N. Sensory difference tests: Some rethinking concerning the general rule that more sensitive tests use fewer stimuli. *Lebensm.-Wiss. u. Technol.* 19: 93-94;1986.
15. Pokorny J, Marcín A, Davídek J. Comparison of the efficiency of triangle and tetrad tests for discriminatory sensory analysis of food. *Die Nahrung* 25: 561-564;1981.
16. Wasserman AE, Talley F. A sample bias in the evaluation of smoked frankfurters by the triangle test. *J. Food Sci.* 34: 99-100;1969.
17. Thurstone LL. A law of comparative judgment. *Psychol. Rev.* 34: 273-286;1927<sup>a</sup>.
18. Thurstone LL. Psychophysical analysis. *Am. J. Psychol.* 38: 368-389;1927<sup>b</sup>.

19. Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, NY, USA. 1966.
20. Macmillan, N.A. and Creelman, C.D. *Detection Theory: A User's Guide*. Cambridge University Press, Cambridge, UK. 1991.
21. Ura S. Pair, triangle and dúo-trio test. *Reports of Statistical Application Res. Japanese Union of Scientists and Engineers* 7: 107-119;1960.
22. Ennis DM. Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *J. Exptl. Psychol. Gen.* 117: 408-411;1988<sup>a</sup>.
23. Ennis DM. Multivariate sensory analysis. *Food Technol.* 42: 118-122;1988b.
24. Ennis DM. Relative power of difference testing methods in sensory evaluation. *Food Technol.* 44: 114-118;1990.
25. Ennis DM. Modeling similarity and identification when there are momentary fluctuations in psychological magnitude. pp. 279-298. In: *Multidimensional models of perception and cognition*. Ashby FG (ed). Lawrence Erlbaum Associates, Hillsdale, NJ, USA. 1992.
26. Ennis DM, Mullen K. The effect of dimensionality on results from the triangular method. *Chem. Senses* 10: 605-608; 1985.
27. Ennis DM, Mullen K. A multivariate model for discrimination methods. *J. Math. Psychol.* 30: 206-219;1986<sup>a</sup>.
28. Ennis DM, Mullen K. Theoretical aspects of sensory discrimination. *Chem. Senses* 11: 513-522;1986b.
29. Ennis DM, Mullen K. A general probabilistic model for triad discrimination, preferential choice, and two-alternative identification. pp. 115-122. In: *Multidimensional models of perception and cognition*. Ashby FG (ed). Lawrence Erlbaum Associates, Hillsdale, NJ, USA. 1992<sup>a</sup>.
30. Ennis DM, Mullen K. Probabilistic psychophysics with noisy stimuli. *Math. Soc. Sci.* 23: 221-234;1992b.
31. Ennis DM, O'Mahony M. Probabilistic models for sequential taste effects in triadic choice. *J. Exptl. Psychol.-Human Perception and Performance* 21: 1088-1097;1995.
32. Ennis DM, Mullen K, Frijters JER. Variants of the method of triads: Unidimensional Thurstonian models. *Brit. J. Math. Stat. Psychol.* 41: 25-36;1988<sup>a</sup>.
33. Ennis DM, Palen JJ, Mullen K. A multidimensional stochastic theory of similarity. *J. Math. Psychol.* 32: 449-465;1988b.
34. Ennis JM, Ennis DM, Yip D, O'Mahony M. Thurstonian models for variants of the method of tetrads. *Brit. J. Math. Stat. Psychol.* 51, 205-215;1998.
35. Frijters JER. The paradox of discriminatory nondiscriminators resolved. *Chem. Senses Flavor* 4: 355-358;1979<sup>a</sup>.
36. Frijters, J.E.R. Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced-choice signal detection theory models. *Brit. J. Math. Stat. Psychol.* 32: 229-241;1979b.
37. Frijters JER. Three-stimulus procedures in olfactory psychophysics: An experimental comparison of Thurstone-Ura and three-alternative forced-choice models of Signal Detection theory. *Percept. Psychophys.* 28: 390-397;1980.
38. Frijters JER. The excitatory state in the triangular constant method. *Psychometrika* 46: 219-222;1981<sup>a</sup>.
39. Frijters, J.E. An olfactory investigation of the compatibility of oddity instructions with the design of a 3-AFC Signal Detection task. *Acta Psychol.* 49: 1-16;1981b.
40. Kapenga JA, de Doncker E, Mullen K, Ennis DM. The integration of the multivariate normal density function for the triangular method. pp. 321-328. In *Numerical Integration: Recent Developments, Software, and Applications*. Keast P, Fairweather G (ed). D. Reidel Publ. Co., Dordrecht, Holland.1987.
41. Mosteller F. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika* 16: 3-9;1951<sup>a</sup>.
42. Mosteller F. Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. *Psychometrika* 16: 203-206;1951b.
43. Mosteller F. Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika* 16: 207-218;1951c.
44. Mullen K, Ennis DM. Mathematical formulation of multivariate Euclidean models for discrimination methods. *Psychometrika* 52: 235-249;1987.
45. Mullen K, Ennis DM. A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika* 56: 69-75;1991.
46. Mullen K, Ennis DM, de Doncker E, Kapenga JA. Models for the dúo-trio and triangular methods. *Biometrics* 44: 1169-1175;1988.
47. Vessereau A. Les méthodes statistiques appliquées au test des caracteres organoleptiques. *Ann. Nutr.* 19: 103-140;1965.
48. Elliott PB. Tables of  $d'$ . pp. 651-684. In: *Signal Detection and Recognition by Human Observers; Contemporary Readings*. Swets JA (ed). John Wiley and Sons, NY, USA;1964.
49. Ennis DM. The power of sensory discrimination methods. *J. Sens. Stud.* 8: 353-370;1993.
50. Frijters JE. Expanded tables for conversion of a proportion of correct responses ( $P_c$ ) to the measure of sensory difference ( $d'$ ) for the triangular method and the 3-alternative forced choice procedure. *J. Food Sci.* 47: 139-143;1982.
51. Frijters JER, Kooistra A, Vereijken PFG. Tables of  $d'$  for the triangular method and the 3-AFC Signal Detection procedure. *Percept. Psychophys.* 27: 176-178;1980.
52. Hacker MJ, Ratcliff R. A revised table of  $d'$  for M-alternative forced choice. *Percept. Psychophys.* 26: 168-170;1979.
53. O'Mahony M. Understanding discrimination tests: A user-friendly treatment of response bias, rating and ranking R-index tests and their relationship to signal detection. *J. Sens. Stud.* 7: 1-47;1992.
54. O'Mahony M. Who told you the triangle test was simple? *Food Qual. Pref.* 6:227-238;1995.
55. doxes and cognitive strategies. *J. Sens. Stud.* 9: 247-272;1994.
56. O'Mahony M, Rousseau B. Discrimination testing: A few ideas, old and new. *Food Qual. Pref.* 14: 157-164;2002.
57. Abrahams H, Krakauer D, Dallenbach KM. Gustatory adaptation to salt. *Am. J. Psychol.* 49: 462-469 (1937).
58. Yip DH. Triadic and tetradic taste discrimination testing: Thurstonian and sequential effects. MS Thesis, University of California, Davis, CA, USA;1996.

59. Delwiche J, O'Mahony M. Flavour discrimination - An extension of Thurstonian paradoxes to the tetrad method. *Food Qual. Pref.* 7: 1-5;1996.
60. Masuoka S, Hatjopoulos D, O'Mahony M. Beer bitterness detection: Testing Thurstonian and Sequential Sensitivity Analysis models for triad and tetrad methods. *J. Sens. Stud.* 10: 295-306;1995.
61. Rousseau B, O'Mahony M. Sensory difference tests: Thurstonian and SSA predictions for vanilla flavored yogurts. *J. Sens. Stud.* 12: 127-146;1997.
62. Stillman JA. Response selection, sensitivity, and taste-test performance. *Percept. Psychophys.* 54: 190-194;1993.
63. Tedja S, Nonaka R, Ennis DM, O'Mahony M. Triadic discrimination testing -Refinement of Thurstonian and Sequential Sensitivity Analysis approaches. *Chem. Senses* 19: 279-30;1994.
64. Bi J, Ennis DM, O'Mahony M. How to estimate and use the variance of  $d'$  from difference tests. *J. Sens. Stud.* 12: 87-104;1997.
65. Dessirier J-M, O'Mahony M. Comparison of  $d'$  values for the 2-AFC (paired comparison) and 3-AFCdiscrimination methods: Thurstonian models, sequential sensitivity analysis and power. *Food Qual. Pref.* 10: 51-58;1999.
66. Dessirier J-M, Sieffermann J-M, O'Mahony M. Taste discrimination by the 3-AFC method: testing sensitivity predictions regarding particular tasting sequences based on the sequential sensitivity analysis model. *J. Sens. Stud.* 14: 271-287;1999.
67. O'Mahony M. Salt taste adaptation: the psychophysical effects of adapting solutions and residual stimuli from prior tastings on the taste of sodium chloride. *Perception* 8: 441-476;1979.
68. O'Mahony M, Goldstein LR. Effectiveness of sensory difference tests: Sequential Sensitivity Analysis for liquid food stimuli. *J. Food Sci.*, 51: 1550-1553;1986.
69. O'Mahony M, Goldstein LR. Tasting successive salt and water stimuli: the roles of adaptation, variability in physical signal strength, learning, supra- and subadapting signal detectability. *Chem. Senses* 12: 425-436;1987.
70. O'Mahony M, Odibert N. A comparison of sensory difference procedures: sequential Sensitivity Analysis and aspects of taste adaptation. *J. Food Sci.* 50:1055-1058;1985.
71. Rousseau B, Meyer A, O'Mahony M. Power and sensitivity of the same-different test: Comparison with triangle and dúo-trio methods. *J. Sens. Stud.* 13: 149-173;1998.
72. Rousseau B, Rogeaux M, O'Mahony M. Mustard discrimination by same-different and triangle tests: aspects of irritation, memory and  $\delta$  criteria. *Food Qual. Pref.* 10:173-184;1999.
73. Vié A, O'Mahony M. Triangular difference testing: Refinements to Sequential Sensitivity Analysis for predictions for individual triads. *J. Sens. Stud.* 4: 87-103;1989.
74. Avancini de Almeida TC, Cubero E, O'Mahony M. Same-different discrimination tests with interstimulus delays up to one day. *J. Sens. Stud.* 14: 1-18;1999.
75. Cubero E, Avancini de Almeida TC, O'Mahony M. Cognitive aspects of difference testing: Memory and interstimulus delay. *J. Sens. Stud.* 10: 307-324;1995.
76. Rousseau B, Stroh S, O'Mahony M. Investigating more powerful discrimination tests with consumers: effects of memory and response bias. *Food Qual. Pref.* 13:39-45;2002.

Recibido: 10-12-2008

Aceptado: 12-05-2009