

UNIVERSIDAD DE COSTA RICA  
FACULTAD DE CIENCIAS AGROALIMENTARIAS  
ESCUELA DE TECNOLOGIA DE ALIMENTOS  
LICENCIATURA EN INGENIERIA DE ALIMENTOS

Proyecto final de graduación presentado a la Escuela de Tecnología de Alimentos  
para optar por el grado de Licenciatura en Ingeniería de Alimentos

Comparación del análisis tradicional de la escala hedónica de 9 puntos, que da  
resultados sesgados, con un nuevo tipo de análisis basado en la teoría de  
detección de señales llamado el índice R para ordenamiento

---

**Fiorela Jara Solís**

**B23374**

Ciudad Universitaria Rodrigo Facio  
San José, Costa Rica  
Junio, 2018

## **TRIBUNAL EXAMINADOR**

Proyecto de graduación presentado a la escuela de Tecnología de Alimentos como requisito parcial para optar por el grado de Licenciatura en Ingeniería de Alimentos.

Elaborado por: Fiorela Jara Solís

Aprobado por:

---

**M.Sc. Ana Isabel Incer González**  
**Presidenta del Tribunal**

---

**Ph.D. Elba Cubero Castillo**  
**Directora del proyecto**

---

**MGA. Yorleny Araya Quesada**  
**Asesora del proyecto**

---

**M.Sc. Marcia Cordero García**  
**Asesora del proyecto**

---

**Licda. Sandra Calderón Villaplana**  
**Profesora designada**

## **DEDICATORIA**

Dedicado a Dios, por conducirme en la toma de decisiones; a papi, mami y Naty por construir la persona que soy; a Les por estar a mi lado cuando más la necesitaba; a las profes Elba, Yorle y Marcia por el apoyo y las motivaciones.

## **AGRADECIMIENTOS**

Gracias primeramente a Dios por permitirme cumplir metas. A papi por ayudarme a vivir la vida, aunque sea con plato chico y piñón grande, a mami por construir lo que soy, enseñarme que la familia es lo primero y que la vida es eso que va más allá de lo académico; y a Naty por soportarme en los momentos más difíciles. A Les por enseñarme que las situaciones de la vida no son sencillas, pero si las cargamos juntas son menos pesadas; por cada regaño, por cada apoyo y por a pesar de todo, seguir aquí.

A Ronital y las chicas: Caro, Esther, Vivi, Megan, Steph G, por tener siempre una palabra de motivación, por creer en mí, ayudarme hasta en las tareas más difíciles y tediosas, por estar en mis decisiones más importantes, por aguantar mis berrinches, aceptar mis críticas y celebrar mis triunfos.

A Elena, Ari, Steph A, Maricruz, Dani, Monse, Rebe, Xime, Rody, Camilo, Mar, Lali, Smith, Carrillo, Vale, Mariel, Jhos, Mario y todos los amigos y compañeros, por sus bromas, consejos y motivaciones. Nunca los olvidaré. A Naty Lau y a Vale B. por hacer la U muy divertida, por los consejos, por cada risa y preocupación que vivimos juntas.

A todos esos amigos que se las ingeniaban para darme un abrazo, los que esperaban el día de mi cumpleaños o el primer día de clases (Ari). Al mejor asesor financiero, Giova, por las enseñanzas y consejos que nunca olvidaré. ¡Son solo negocios! A Luisito por impulsarme a dar ese paso en el camino.

A la profe Elba por confiar en mí, por todas las enseñanzas y por sacar un espacio para mis dudas; a la profe Yorle por cada consejo y por estar dispuesta siempre a ayudar; y a la profe Marcia por la motivación y la ayuda en la obtención de resultados. A la profe Ile y a la profe Marcia por tener las palabras correctas en el momento que más lo necesitaba. Al personal de la escuela por toda su ayuda, Nela, Taty, Doña Rocio, Luis David y todos los que formaron parte de mi formación académica y personal. A todas y cada una de esas personas que con solo preguntar por la tesis me impulsaban a escribir.

## ÍNDICE GENERAL

TRIBUNAL EXAMINADOR .....	i
DEDICATORIA.....	ii
AGRADECIMIENTOS.....	iii
ÍNDICE GENERAL .....	iv
ÍNDICE DE FIGURAS.....	vii
ÍNDICE DE CUADROS.....	viii
RESUMEN .....	x
1. JUSTIFICACIÓN .....	1
2. OBJETIVOS.....	6
2.1. Objetivo General .....	6
2.2. Objetivos específicos.....	6
3. MARCO TEÓRICO .....	7
3.1. Análisis sensorial .....	7
3.2. Pruebas orientadas al consumidor.....	8
3.3. Tipos de escalas .....	9
3.3.1. Escalas nominales.....	9
3.3.2. Escalas ordinales.....	9
3.3.3. Escalas proporcionales.....	9
3.3.4. Escalas de intervalo .....	9
3.4. Escalas hedónicas .....	10
3.4.1. Escala hedónica de 9 puntos.....	11
3.5. Estrategias cognitivas utilizadas en las pruebas de consumidores con escala hedónica .....	13
3.5.1. Modelo Absoluto de Zwislocki .....	13
3.5.2. Modelo relativo de Mellers .....	13
3.6. Traducciones de la escala hedónica de 9 puntos al español .....	14
3.7. Limitaciones de la escala hedónica de 9 puntos .....	16
3.7.1. Uso de la escala, significado y orden de las frases de las categorías.....	16
3.7.2. Espaciamiento psicológico de las frases, utilización de números y palabras. .....	16
3.7.3. Número limitado de categorías.....	17

3.7.4. Efecto de tendencia central.....	18
3.7.5. Análisis estadístico.....	19
3.8. Análisis de datos.....	19
3.8.1. Análisis de varianza (ANDEVA).....	19
3.8.2. Índice R.....	20
3.8.3. Interpretación del análisis de resultados de la escala hedónica de 9 puntos.....	23
4. MATERIALES Y MÉTODOS.....	24
4.1. Localización.....	24
4.2. Escogencia de productos: prueba preliminar.....	24
4.3. Traducción de términos de la escala.....	26
4.4. Evaluación del agrado de los productos usando menos muestras que categorías.....	28
4.4.1. Escala.....	28
4.4.2. Productos.....	28
4.4.3. Evaluación.....	29
4.4.3.1. Consumidores.....	29
4.5. Evaluación del agrado de los productos usando más muestras que categorías.....	29
4.5.1. Escala.....	29
4.5.2. Productos.....	29
4.5.3. Evaluación.....	30
4.5.3.1. Consumidores.....	30
4.6. Análisis.....	31
4.6.1. Análisis de datos tradicional.....	31
4.6.2. Análisis mediante índice R de ordenamiento.....	31
4.6.2.1. $R_{JB}$ .....	32
4.6.2.1.1. Menos muestras que categorías (5 gomitas).....	32
4.6.2.1.2. Más muestras que categorías (13 gomitas).....	33
4.6.2.2. $R_{MAT}$ .....	35
4.6.2.2.1. Menos muestras que categorías (5 gomitas).....	35
4.6.2.2.2. Más muestras que categorías (13 gomitas).....	36
4.6.3. Comparación de resultados.....	38
4.6.4. Comparación de los índices $R_{JB}$ y $R_{MAT}$ para dos tamaños de muestras.....	38

5.	RESULTADOS Y DISCUSIÓN .....	39
5.1.	Escala hedónica de 9 puntos traducida al español .....	39
5.2.	Análisis estadístico paramétrico y no paramétrico de la evaluación de más muestras que categorías .....	40
5.2.1.	Análisis tradicional (paramétrico) .....	40
5.2.2.	Índice R para ordenamiento.....	43
5.2.2.1.	$R_{MAT}$ .....	43
5.2.2.2.	$R_{JB}$ .....	44
5.3.	Análisis estadístico paramétrico y no paramétrico de menos muestras que categorías.....	46
5.3.1.	Análisis tradicional.....	46
5.3.2.	Índice R para ordenamiento.....	47
5.3.2.1.	$R_{MAT}$ .....	47
5.3.2.2.	$R_{JB}$ .....	48
5.4.	Comparación del índice $R_{JB}$ y $R_{MAT}$ para dos tamaños de muestras .....	49
6.	CONCLUSIONES .....	51
7.	RECOMENDACIONES.....	52
8.	BIBLIOGRAFÍA .....	53
9.	ANEXOS.....	58
9.1.	Hoja de recolección de datos para análisis de cinco muestras.....	58
9.2.	Matriz para obtención del índice $R_{MAT}$ para cinco muestras.....	59
9.3.	Matriz para obtención del índice $R_{MAT}$ para trece muestras .....	60
9.4.	Muestra de cálculos del índice R para ordenamiento.....	61
9.5.	Herramientas para la construcción de la escala hedónica de 9 puntos traducida al español. ....	65

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Escala hedónica de 9 puntos utilizada en las evaluaciones sensoriales de agrado de gomitas. ....	28
<b>Figura 2.</b> Tarjetas utilizadas en la construcción de la escala hedónica de 9 puntos traducida al español. ....	65
<b>Figura 3.</b> Ejemplo de las categorías elaboradas por una muestra de la población costarricense. ....	65

## ÍNDICE DE CUADROS

<b>Cuadro I.</b> Categorías de la escala hedónica de 9 puntos en idioma inglés. ....	12
<b>Cuadro II.</b> Categorías de la escala hedónica de 9 puntos en idioma inglés y cuatro traducciones al idioma español. ....	15
<b>Cuadro III.</b> Muestras de gomitas analizadas en la escogencia del producto a utilizar; con su respectiva marca, clasificación y observaciones. ....	25
<b>Cuadro V.</b> Muestras de gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos. ....	28
<b>Cuadro VI.</b> Muestras de gomitas utilizadas en la evaluación de más muestras que categorías mediante la escala hedónica de 9 puntos. ....	30
<b>Cuadro VII.</b> Valor numérico asignado a cada una de las categorías de la escala hedónica de nueve (Wichchukit & O'Mahony, 2015). ....	31
<b>Cuadro VIII.</b> Muestras de gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos y su código respectivo. ....	32
<b>Cuadro IX.</b> Clasificación de las categorías del ordenamiento de las gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos por cada juez, donde 1º fue la de mayor agrado y los valores mayores menos agradables. ....	32
<b>Cuadro X.</b> Muestras de gomitas utilizadas en la evaluación de más muestras que categorías mediante la escala hedónica de 9 puntos y su código respectivo. ....	33
<b>Cuadro XI.</b> Clasificación de las categorías del ordenamiento de las gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos por cada juez. ....	34
<b>Cuadro XII.</b> Clasificación de los cinco productos evaluados mediante la escala hedónica de 9 puntos por cada juez. ....	35
<b>Cuadro XIII.</b> Posicionamiento (ordenamiento) de dos de los productos evaluados mediante la escala hedónica de 9 puntos. ....	36
<b>Cuadro XIV.</b> Clasificación por ordenamiento de los trece productos evaluados mediante la escala hedónica de 9 puntos por cada juez. ....	37
<b>Cuadro XV.</b> Posicionamiento (ordenamiento) de dos de los productos evaluados mediante la escala hedónica de 9 puntos. ....	37
<b>Cuadro XVI.</b> Categorías de la escala hedónica de 9 puntos en inglés, traducción directa al español y escala obtenida para la evaluación de los productos. ....	39

<b>Cuadro XVII.</b> Análisis de varianza para determinar diferencias en el agrado de 13 muestras de gomitas mediante la escala hedónica de 9 puntos.....	41
<b>Cuadro XVIII.</b> Resultados de la comparación de medias de las calificaciones dadas por los consumidores a las trece muestras analizadas .....	41
<b>Cuadro XIX.</b> Índice $R_{MAT}$ para ordenamiento obtenido para trece productos mediante la utilización de la escala hedónica de 9 puntos.....	43
<b>Cuadro XX.</b> Índice $R_{JB}$ para ordenamiento obtenido para trece productos mediante la utilización de la escala hedónica de 9 puntos.....	45
<b>Cuadro XXI.</b> Análisis de varianza para determinar diferencias en el agrado de 5 muestras de gomitas mediante la escala hedónica de 9 puntos.....	46
<b>Cuadro XXII.</b> Resultados de la comparación de medias de las calificaciones dadas por los consumidores a las cinco muestras analizadas.....	46
<b>Cuadro XXIII.</b> Índice $R$ para ordenamiento ( $R_{MAT}$ ) obtenido para cinco productos mediante la utilización de la escala hedónica de 9 puntos.....	47
<b>Cuadro XXIV.</b> Índice $R_{JB}$ para ordenamiento obtenido para cinco productos mediante la utilización de la escala hedónica de 9 puntos.....	48
<b>Cuadro XXV.</b> Número de veces que $R_{JB} > R_{MAT}$ y $R_{MAT} > R_{JB}$ a partir de las diferencias significativas detectadas en al menos uno de los índices $R$ para ordenamiento, para dos tamaños de muestras distintas.....	49
<b>Cuadro XXVI.</b> Matriz para la obtención del índice $R_{MAT}$ para un tamaño de muestras menor al número de categorías de la escala hedónica de 9 puntos. ...	59
<b>Cuadro XXVII.</b> Matriz para la obtención del índice $R_{MAT}$ para un tamaño de muestras mayor al número de categorías de la escala hedónica de 9 puntos.....	60
<b>Cuadro XXVIII.</b> Posición de cada muestra de la evaluación de 5 productos mediante la escala hedónica de 9 puntos usando 20 jueces. ....	61
<b>Cuadro XXIX.</b> Número de veces que cada producto se encuentra en determinada posición, obtenido mediante la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos. ....	63
<b>Cuadro XXX.</b> Comparación de las diferencias encontradas para las trece muestras analizadas con estadística paramétrica y los índices $R$ . ....	66
<b>Cuadro XXXI.</b> Comparación de las diferencias encontradas para las cinco muestras analizadas con estadística paramétrica y los índices $R$ . ....	66

## RESUMEN

Jara Solís, Fiorela

Comparación del análisis tradicional de la escala hedónica de 9 puntos, que da resultados sesgados, con un nuevo tipo de análisis basado en la teoría de detección de señales llamado el índice R para ordenamiento

Tesis en Ingeniería de Alimentos, San José, CR:

F Jara S., 2018

67 pag.: 3 il.- 58 ref.

El objetivo de la investigación fue comparar el análisis tradicional de la escala hedónica de 9 puntos con un nuevo tipo de análisis basado en la teoría de detección de señales, llamado el índice R para ordenamiento, usando dos tamaños de muestras para la corrección de un análisis de datos sesgado.

Se construyó la escala hedónica de 9 puntos según el lenguaje del costarricense con 50 personas, a las cuales se les permitía construir hasta tres escalas y utilizar traducciones de la escala presentadas o incluir sus propias frases. Se analizaron dos tamaños de muestras de gomitas, uno mayor y otro menor que el número de categorías de la escala (13 y 5) con ayuda de la escala construida.

Se realizaron dos tipos de análisis estadístico de los resultados, el análisis paramétrico tradicional (ANDEVA) y el análisis no paramétrico (R-Index). Se obtuvo el índice R de dos maneras distintas  $R_{JB}$  y  $R_{MAT}$ , los cuales fueron comparados con la finalidad de determinar si existe diferencia en la detección de variaciones de agrado.

La escala obtenida no coincidió con la traducción directa de la escala hedónica en español, debido a que se presentaron variaciones en los extremos de la escala. La estadística no paramétrica fue más sensible que la paramétrica para detectar diferencias, es decir, encontró mayor cantidad de diferencias entre los productos al utilizar un tamaño de muestras mayor y uno menor que el número de categorías de la escala.

Se evidenció que el índice  $R_{JB}$  permitió detectar mayores diferencias en mayor número de comparaciones que el índice  $R_{MAT}$ , para dos tamaños de muestras distintas (5 y 13)

Ph.D. Elba Cubero Castillo

Escuela de Tecnología de Alimentos

## 1. JUSTIFICACIÓN

El crecimiento y desarrollo del análisis sensorial ha tenido un impacto significativo en el ámbito alimentario; lo cual se asocia al uso de las pruebas de agrado, que están orientadas al consumidor. En la industria alimentaria se ha incrementado la utilización de estas pruebas debido a la relevancia que poseen en el diseño de un producto nuevo o la modificación de un producto establecido en el mercado, al dar información sobre el agrado y preferencias del consumidor (Ramírez, 2012).

En el desarrollo de pruebas de consumidor se hace uso de diversos tipos de escalas, como lo son las escalas nominales, escalas ordinales, escalas proporcionales y escalas de intervalo. La primera de éstas se basa en escalas donde las variables de estudio son cualitativas a pesar de estar asociadas a números, mientras que la ordinal se basa en la utilización de números para conocer el orden de preferencia de los productos analizados (González *et al.*, 2014). Las escalas proporcionales y de intervalo se fundamentan en que los números expresan cantidades reales que generan información sobre las diferencias del grado de aceptación de los productos. La diferencia entre éstas consiste en que en la primera el cero es real y representa la ausencia del atributo, mientras que en la de intervalo el cero es únicamente una referencia, más no indica la ausencia del atributo (González *et al.*, 2014).

En el campo de la industria alimentaria se emplea generalmente la escala hedónica de 9 puntos para la realización de las pruebas de aceptación del consumidor, caracterizada por poseer un valor central neutro con cuatro categorías positivas y cuatro negativas (Lim, 2011). Esta escala se clasifica como una escala verbal de categorías en la cual se asigna a cada una de éstas un valor numérico que permite el análisis paramétrico, en el cual se establecen espacios iguales entre las frases (Ramírez, 2012). Por lo anterior, se le atribuye que pertenece al tipo de escala de intervalo; no obstante, se ha argumentado que la escala hedónica de 9 puntos debe ser preferentemente utilizada como escala ordinal en contraposición a la escala de intervalo (Nicolas *et al.*, 2010).

En la utilización de la escala hedónica de 9 puntos se debe tener un extremo cuidado en las frases utilizadas con el fin de que éstas sean claras y graduales

para cada uno de los analistas (González *et al.*, 2014). Se sugiere lo anterior ya que inicialmente la utilización de la escala hedónica de 9 puntos se daba en países de habla inglesa; no obstante, se ha expandido a naciones de habla hispana (Curia *et al.*, 2001).

Se ha demostrado la variación presentada en diferentes culturas en cuanto a la apreciación de las frases generadas como traducción directa de la escala hedónica de 9 puntos utilizada en Estados Unidos, donde se ha puesto en evidencia que los consumidores asiáticos utilizan menos categorías en dicha escala. Lo anterior se asocia a que estos consumidores se desarrollaron en una cultura donde no se expresan aspectos negativos, así como el no dar más valor a una muestra sobre otra (Yeh *et al.*, 1998). Sin embargo, no se consideró que las diferencias en el uso de la escala se debieran a la ambigüedad generada al traducir las frases (Curia *et al.*, 2001).

En la traducción al español de las categorías utilizadas en el idioma inglés se ha evidenciado ambigüedad en el significado de las frases. Según Curia *et al.* (2001) en un estudio realizado con pobladores argentinos se demostró que un 30% de los analistas, al utilizar la escala traducida especialmente para niños, generaron un ordenamiento de las categorías diferente al original. Lo anterior respalda lo mencionado sobre la importancia de realizar un estudio exhaustivo sobre la traducción de la frase, el entendimiento que la población estudiada tiene sobre el uso de dicha escala y ordenamiento que los analistas realizan de éstas.

La escala hedónica de 9 puntos genera dificultad para que el panelista exprese sus percepciones sensoriales a causa de presentar un número limitado de categorías (Lim, 2011). Por lo anterior, al utilizar un tamaño de muestras mayor a las categorías existentes en la escala se da la presencia de sesgo en los resultados obtenidos dado que el panelista contaba con menos categorías de las que pudiera necesitar. Contrario a esto, al poseer un número de muestras menor a las categorías presentadas se logra que el analista pueda ubicar el producto en el sitio que considera, clasificando uno o dos productos en una misma categoría. Lo anterior respalda la necesidad de realizar estudios con diferentes números de muestras con el fin de conocer las variaciones presentadas por este factor (Wichchukit & O'Mahony, 2015).

Esta escala se caracteriza por producir un efecto de tendencia central, en la cual los panelistas evaden la utilización de las categorías extremas y reducen la

escala a siete puntos, disminuyendo su capacidad para detectar las diferencias entre muestras. Este efecto se genera por el pequeño número de categorías que dispone el analista, lo cual hace que prefiera la utilización de las categorías centrales en comparación con las externas, con el fin de reservar las últimas (Lim, 2011).

A lo anterior también se le conoce, en el área de la psicología, con el nombre de efecto final de la escala; y se define como la mayor dificultad que posee el analista para pasar de la categoría ocho a nueve en comparación con pasar de la cuarta a la quinta categoría. Esto se genera a causa de que, como se ha mencionado, el espaciamiento entre las categorías no es la misma cognitivamente hablando y, por tanto, se descarta que ésta sea una escala de intervalo (Wichchukit & O'Mahony, 2015).

En el análisis de los resultados obtenidos mediante la escala hedónica de 9 puntos se hace uso de estadística paramétrica o de estadística no paramétrica, lo que está definido por el tratamiento que se les da a los resultados obtenidos (Lim, 2011). Si el valor se clasifica como categórico y discreto se hace uso de la estadística no paramétrica, mientras que si se utilizan como datos continuos se hace uso de la estadística paramétrica (Jaeger & Cardello, 2009).

En el análisis tradicional de la escala hedónica de 9 puntos se atribuye a cada categoría un número con el fin de utilizar estadística paramétrica (Nicolas *et al.*, 2010); sin embargo, se debe analizar la equidistancia en las frases utilizadas en el idioma en español para la utilización de este tipo de estadística en el análisis a realizar (Curia *et al.*, 2001).

La información obtenida mediante el uso de la escala hedónica de nueve puntos no cumple con los supuestos estadísticos de normalidad, homoscedasticidad y aditividad, requeridos en los modelos de análisis de varianza (Villanueva *et al.*, 2005); no obstante, es el método más utilizado para el análisis de los datos obtenidos (Lim, 2011). Por lo anterior, se establece que el uso de la escala no es el causante de los problemas, sino el análisis estadístico que se realiza de los datos obtenidos a partir de ésta, al utilizarlos como si fueran generados por instrumentos calibrados y no por personas con diferentes capacidades cognitivas (Wichchukit & O'Mahony, 2015).

Se ha demostrado que al asumir que existe una diferencia de una unidad numérica entre cada frase se está incurriendo en un sesgo ya que se ha

evidenciado que al utilizar la escala hedónica de 9 puntos únicamente con palabras contra dicha escala en presencia de números se obtienen diferentes resultados (Nicolas *et al.*, 2010), y esto se atribuye a que las distancias psicológicas entre las frases no son iguales a las presentadas en los valores numéricos (Lim, 2011). Lo anterior se asocia a que la estrategia cognitiva que utiliza el analista es diferente en ambos casos, dado que la diferencia entre frases no es fija (Wichchukit & O'Mahony, 2015).

Las limitaciones presentadas por la escala hedónica de 9 puntos en cuanto a su poder discriminativo y el análisis estadístico paramétrico han generado que se propongan diversas alternativas para la solución de dichos inconvenientes (Jaeger & Cardello, 2009), entre los cuales se encuentran variaciones en el análisis de los datos y en escalas alternativas como lo son la escala hedónica híbrida, la escala auto ajustable y el ordenamiento (Villanueva *et al.*, 2005; Varela *et al.*, 2014).

En estudios anteriores se ha demostrado que la validez de la estadística paramétrica es cuestionable; a pesar de ello, se establece que dicho análisis es necesario y se justifica su utilización por motivos prácticos hasta que se disponga de una técnica apropiada para tal fin (Wichchukit & O'Mahony, 2014). Una alternativa para el análisis de la escala hedónica de 9 puntos es la utilización del índice R para ordenamiento, el cual convierte la respuesta de la escala hedónica de 9 puntos a números ordinales y hace uso del análisis de detección de señales para obtener datos no sesgados.

Este método de detección de señales obtiene la misma información que con la escala hedónica de forma numérica sin necesidad de recurrir a ésta (Wichchukit & O'Mahony, 2014). El índice R para ordenamiento es una probabilidad estimada para identificar un estímulo objetivo (señal) cuando se presenta un segundo estímulo (ruido), es decir, distinguir un estímulo del no estímulo. Lo anterior le permite ser utilizado como alternativa en procedimientos de pruebas de diferencias de elección forzada y en pruebas de escala (Lee & Van Hout, 2009).

Entre las ventajas de este índice se encuentra que no solamente indica que dos productos son diferentes, sino que da un tamaño de diferencia (Varela *et al.*, 2014). De igual manera está libre de los sesgos de respuesta o variación de criterio y se conoce como un elemento cognitivo en la toma de decisiones acerca

de si una sensación indica una diferencia o no; esto debido a ser una medida de detección de señales (Lee & Van Hout, 2009).

El cálculo del índice R se utiliza bajo un enfoque alternativo, el cual consiste en la teoría de detección de señales, donde se le solicita al juez si está seguro o no de su juicio (Lee & Ven Hout, 2009).

Cuando se utilizan muchas muestras, el uso del análisis mediante el índice R para ordenamiento presenta limitaciones, ya que pueden encontrarse dos o más muestras en la misma categoría. Estas limitaciones pueden ser solucionadas utilizando una categorización previa, utilizando escala hedónica de 9 puntos (Wichchukit & O`Mahony, 2014), y posteriormente se clasifican las muestras que se ubicaron en la misma categoría. Con lo anterior se genera una clasificación general, asegurando que el último estímulo de la primera categoría esté por encima del primer estímulo de la segunda categoría (Lee & Hout, 2009).

La diferencia entre el análisis sensorial tradicional y la teoría de detección de señales recae en que esta última se basa en la comprensión de que la percepción de un alimento es variable, que los cambios en la boca y las fluctuaciones en el sistema nervioso causan cambios leves en la intensidad que se percibe en un producto luego de una degustación repetida (Lee & Van Hout, 2009) y, por lo tanto, la evaluación de la misma muestra puede variar en mediciones repetidas.

Basado en lo expuesto anteriormente se establece la necesidad de comparar el análisis tradicional de la escala hedónica de 9 puntos, que asume el uso de las frases como número igualmente espaciados, que da resultados sesgados, con un nuevo tipo de análisis basado en la teoría de detección de señales llamado el índice R de ordenamiento, donde se tomará en cuenta que pueden haber más de una muestra en una misma categoría y que las frases están igualmente espaciadas.

## 2. OBJETIVOS

### 2.1. Objetivo General

Comparar el análisis tradicional de la escala hedónica de 9 puntos, traducida al español, con un nuevo tipo de análisis basado en la teoría de detección de señales, llamado el índice R para ordenamiento, usando un tamaño de muestra mayor y uno menor que el número de categorías para la corrección de un análisis de datos sesgado.

### 2.2. Objetivos específicos

1. Adaptar las traducciones y el orden de las categorías de la escala hedónica de 9 puntos que genere el mayor entendimiento por parte de una muestra de la población costarricense.
2. Comparar los resultados obtenidos con el análisis tradicional de la escala hedónica de 9 puntos con el índice R para ordenamiento cuando se evalúa una cantidad de muestras mayor que el número de categorías para la corrección del sesgo del análisis tradicional.
3. Comparar los resultados que se obtienen del análisis tradicional de la escala hedónica de 9 puntos respecto al obtenido con el índice R para ordenamiento para dos tamaños de muestras, para la definición del efecto del número de muestras evaluadas sobre el sesgo del análisis tradicional.
4. Comparar los resultados de los índices R para ordenamiento, índice  $R_{JB}$  y  $R_{MAT}$ , para dos tamaños de muestras.

### 3. MARCO TEÓRICO

#### 3.1. Análisis sensorial

El análisis sensorial es una disciplina científica que permite la medición, análisis e interpretación de la reacción ante las características de alimentos que son percibidas a través de los sentidos (Ibáñez & Barcina, 2001; Ramírez, 2012).

Los análisis utilizados se basan principalmente en métodos cuantitativos debido a que debe cuantificar las sensaciones percibidas por los sentidos otorgándoles un valor numérico (Clark *et al.*, 2009; Ibáñez & Barcina, 2001). Las técnicas del análisis sensorial se basan en las necesidades y el objetivo que se persigue, con lo cual se han clasificado en dos grandes grupos (González *et al.*, 2014):

1. Pruebas analíticas: Buscan medir las características de un producto y se conocen también como pruebas orientadas al producto. En esta rama se encuentran las pruebas de discriminación y descriptivas; para las cuales el panelista es seleccionado por su capacidad sensorial en las características en estudio, y pueden ser entrenados o no según el requerimiento. De igual manera en estas técnicas el panelista debe dejar de lado sus preferencias y enfocarse en atributos específicos (Lawless & Heymann, 2013).
2. Pruebas afectivas: Conocidas también como pruebas de consumidores que son utilizadas para analizar las preferencias de los consumidores o medir el grado de satisfacción proporcionado por un producto. Los consumidores actúan en forma más integrada al percibir un producto de forma general y no por atributos específicos (Watts *et al.*, 1989; Lawless & Heymann, 2013).

El diseño de una prueba sensorial implica la selección de un método adecuado con participantes apropiados y los análisis estadísticos requeridos acorde al objetivo (Lawless & Heymann, 2013), con la finalidad de tener un importante grado de fiabilidad en las pruebas sensoriales, así como resultados cuantificables y reproducibles (Sancho *et al.*, 1999; Meilgaard *et al.*, 2007).

Esta disciplina se caracteriza por formar parte de las áreas de desarrollo de productos y control de calidad de una empresa, tomando en cuenta materias primas, productos en proceso y productos terminados (Sancho Valls *et al.*, 1999; Lee & O'Mahony, 2005). Las pruebas sensoriales aportan a la industria

alimentaria, mediante la caracterización hedónica de productos, la comparación con los alimentos competidores del mercado, el establecimiento de criterios de calidad mediante el desarrollo de un perfil sensorial, el control del proceso de fabricación, la vigilancia del producto integrando aspectos como homogeneidad, la medición de la influencia del almacenamiento en un producto y el desarrollo y lanzamiento de un nuevo producto (Mondino & Ferratto, 2006); todo lo anterior mediante la traducción de las señales, a partir de los sentidos, de panelistas adecuados (Ramírez, 2012; Wichchukit & O'Mahony, 2015).

### **3.2. Pruebas orientadas al consumidor**

Las pruebas orientadas al consumidor permiten la medición del agrado de un producto, así como la determinación del grado de aceptación. Estas pruebas tienen como requisito que los panelistas sean consumidores del producto en estudio (Espinosa *et al.*, 1998; Ramírez, 2012) y no se puede utilizar un panel entrenado debido a que una vez que una persona ha sido entrenada para identificar y cuantificar atributos deja de ser consumidor típico (Clark *et al.*, 2009).

Estas pruebas pueden ser de carácter cualitativo o cuantitativo, según sea el objetivo del análisis. Las primeras de estas se basan en estudios con grupos focales o entrevistas a profundidad, mientras que las pruebas cuantitativas abarcan pruebas que generan resultados que expresan, mediante números, la cantidad de agrado o el nivel de preferencia sobre un producto (Pilgrim & Peryam, 1996).

Las pruebas cuantitativas de consumo tienen tres dimensiones en las investigaciones: 1. Sensorial y hedónica, 2. Conveniencia y facilidad de compra, transporte, conservación, entre otras; y 3. Beneficios del producto relacionados con la salud (Ramírez, 2012). Así mismo pueden realizarse en diferentes lugares como lo son mercados, escuelas, hogares de los consumidores o centros comerciales; donde se recomienda utilizar un panel lo más numeroso posible con la finalidad de indicar aceptabilidad de un producto e identificar defectos en productos (Watts *et al.*, 1989).

Las pruebas afectivas más conocidas se clasifican en dos grandes grupos, las pruebas de aceptación y las pruebas de preferencia (Lawless & Heymann 2010). Las pruebas de preferencia se basan en encontrar el producto que es elegido sobre otro, por parte del consumidor, las cuales son sencillas de realizar

debido a que son muy intuitivas y se requiere de poca explicación al panelista; así mismo pueden ser realizadas por consumidores de diferentes edades, condiciones y lenguaje (González *et al.*, 2014).

Las pruebas de aceptación se conocen también como pruebas hedónicas y se centran en solicitar al consumidor que indique el nivel de agrado de un producto mediante el uso de una escala y, por consiguiente, son conocidas como pruebas de nivel de agrado. Para realizar las pruebas de aceptabilidad se pueden utilizar pruebas de ordenamiento, escalas categorizadas y pruebas de comparación (Ramírez, 2012).

### **3.3. Tipos de escalas**

En las pruebas de consumidor se hace uso de diferentes tipos de escala para obtener resultados, entre estas escalas se encuentran las siguientes:

3.3.1. Escalas nominales: Se utilizan cuando las variables de estudio son cualitativas a pesar de estar asociadas a números (González *et al.*, 2014). Esta escala consiste en una serie de categorías etiquetadas con números o nombres para su identificación; sin embargo, no existe una relación cuantitativa entre los números y no existe un orden lógico de las categorías (Pimentel *et al.*, 2016; Carpenter *et al.*, 2012).

3.3.2. Escalas ordinales: Se basa en la utilización de números, los cuales poseen un orden lógico, para conocer el orden de preferencia de los productos analizados y permite averiguar cuáles son los productos preferidos pero no las diferencias que se presentan entre los productos debido a que las categorías en esta escala no están igualmente espaciadas (González *et al.*, 2014; Pimentel *et al.*, 2016; Carpenter *et al.*, 2012).

3.3.3. Escalas proporcionales: Los números generan información sobre diferencias en el grado de aceptación de diferentes productos asumiendo distancias iguales entre los puntos de la escala (González *et al.*, 2014; Pimentel *et al.*, 2016), son las escalas que más se ajustan al mecanismo de la percepción cuando se evalúan estímulos simples (Ramírez, 2012).

3.3.4. Escalas de intervalo: son similares a las escalas proporcionales; sin embargo, el cero no significa ausencia de atributo, sino es una

referencia (González *et al.*, 2014). Este tipo de escalas son utilizadas para el análisis sensorial con el objetivo de asegurar la validez de los métodos estadísticos paramétricos. Las escalas de intervalo permite el ordenamiento de muestras basándose en la magnitud de una sola característica, así como determinar el grado de diferencia entre muestras (Ramírez, 2012). Las escalas de intervalo tienen el mismo principio de las escalas ordinales, con la característica adicional de que las distancias entre las categorías son iguales (Carpenter *et al.*, 2012).

Las técnicas de escalas implican el uso de números o palabras para expresar la intensidad de los atributos percibidos, para que los datos sean, posteriormente, tratados de forma estadística y se logren extraer conclusiones a partir de ellos (Meilgaard *et al.*, 2007).

### **3.4. Escalas hedónicas**

Las escalas hedónicas son la herramienta de investigación cuantitativa convencional más utilizada en los últimos 60 años para medir consumo (Feng & O'Mahony, 2017). Se caracteriza por poseer diferente número de categorías (Clark *et al.*, 2009) y fue diseñada con el propósito de medir las preferencias alimentarias de los soldados estadounidenses (Peryam & Girardot, 1952; Jones *et al.*, 1955). Las categorías de esta escala se nombran con frases o números que representan diferentes niveles de agrado con continuidad en cuanto a gusto y disgusto (Pimentel *et al.*, 2016).

Esta escala se reconoce como una escala verbal de categorías y se clasifica como una escala ordinal (Carpenter *et al.*, 2012). Sin embargo, para la realización de un análisis paramétrico se asigna a cada una de las categorías un valor numérico (Ramírez, 2012), lo cual no es adecuado estadísticamente ya que los desarrolladores de estas escalas lo confirman, aunque es el método utilizado para obtener resultados numéricos y poder generar valores de agrado (Peryam *et al.*, 1960).

Por lo anterior, las escalas hedónicas se catalogan, por algunos investigadores, como escalas de intervalo; aunque Nicolas *et al.* (2010) mencionan que dichas escalas deben ser preferentemente utilizadas como escalas ordinales en contraposición a las escalas de intervalo. Así mismo,

Pimentel *et al.* (2016) establecen que clasificar las escalas hedónicas como de intervalo es discutible debido a que los intervalos son iguales numéricamente mas no psicológicamente. Debido a este mismo aspecto Wichchukit & O'Mahony (2015) descartan que la escala hedónica de 9 puntos sea una escala de intervalo, ya que resaltan que requiere mayor esfuerzo, psicológicamente, pasar de la octava a la novena categoría en comparación a pasar de la cuarta a la quinta, por el efecto del uso de final de la escala.

Se pueden utilizar diferentes formas y tamaños de escalas hedónicas, siendo las más comunes las de nueve y siete puntos; no obstante, existen, además de éstas, las de 10, 15 y hasta 20 categorías (Clark *et al.*, 2009). La variación del número de categorías es aceptable siempre que no se utilicen menos de cinco (Pilgrim & Peryam, 1996). Jones *et al.* (1955) mostraron que las escalas más largas, hasta nueve intervalos, tendían a ser más discriminatorias que las escalas más cortas. De igual manera Lim (2011) establece que el número de categorías de una escala está relacionado con la cantidad de estímulos que se están evaluando, y entre más estímulos tengan más categorías serán necesarias.

#### **3.4.1. Escala hedónica de 9 puntos**

Esta escala se caracteriza por ser bipolar, es decir, presenta un valor central neutro con cuatro categorías de agrado y cuatro de desagrado (Yeh *et al.*, 1998; Lim, 2011); y es el instrumento más utilizado para las pruebas de agrado (Drake, 2007; Pimentel *et al.*, 2016). Esta escala ha sido traducida a diferentes idiomas; sin embargo, fue primeramente utilizada en el idioma inglés con las siguientes etiquetas que se presentan en el Cuadro I (Curia *et al.*, 2001; D. Peryam & Pilgrim, 1957; Pimentel *et al.*, 2016; Wichchukit & O'Mahony, 2015).

**Cuadro I.** Categorías de la escala hedónica de 9 puntos en idioma inglés.

Dislike extremely
Dislike very much
Dislike moderately
Dislike slightly
Neither like or dislike
Like slightly
Like moderately
Like very much
Like extremely

Se escogieron esas categorías ya que se encontró que entre ellas había la misma distancia psicológica, en el idioma inglés. La escala hedónica de 9 puntos puede ser presentada de forma horizontal o vertical y puede presentar primero el gusto o el disgusto, debido a que no se presentan diferencias en los resultados (Pilgrim & Peryam, 1996; Lawless *et al.*, 2010). Así mismo, con el uso de esta escala se permite asignar la misma categoría a más de una muestra (Ramírez, 2012). A la escala que utiliza únicamente palabras se le denomina escala hedónica de 9 puntos verbal y a la que se basa en números con indicación de la dirección, cuando se le asigna un número a cada frase, se le denomina escala hedónica de 9 puntos numérica (Feng & O'Mahony, 2017).

La escala hedónica de 9 puntos se ha utilizado en una amplia variedad de productos, desde su invención en la década de 1940, ganando un éxito considerable para diferentes industrias (Schutz & Cardello, 2001; Clark *et al.*, 2009). Sin embargo, al igual que con cualquier prueba sensorial, es importante recordar que en situaciones específicas se puede requerir de una escala más especializada que la tradicional escala hedónica de 9 puntos (Clark *et al.*, 2009).

La forma de presentación de las muestras a evaluar con la escala hedónica de 9 puntos es monádica, es decir, un producto por vez; para obtener evaluaciones absolutas y eliminar el efecto de contexto. Lo anterior con la finalidad de simular las condiciones de compra y consumo de productos; en las cuales se prueba un producto por compra o por consumo, y al comparar con otro producto se hace basado en la memoria. Lo ideal es realizar la evaluación de los

productos en días o semanas diferentes, no obstante en la industria es poco práctico (Pimentel *et al.*, 2016).

### **3.5. Estrategias cognitivas utilizadas en las pruebas de consumidores con escala hedónica**

Al utilizar la escala hedónica de 9 puntos se obtienen resultados en forma numérica; no obstante, se establece que la diferencias entre las categorías no son homogéneas (Peryam & Girardot, 1952; Peryam & Pilgrim, 1957) y, por tanto, se presentan dos estrategias cognitivas o reglas de decisión utilizadas por el cerebro: el modelo absoluto de Zwislocki (Zwislocki & Goodman, 1980; Zwislocki, 1983) y el modelo relativo de Mellers (Mellers, 1983a,b).

3.5.1. Modelo Absoluto de Zwislocki: Se basa en la hipótesis de que un estímulo o intensidad a evaluar se compara con un conjunto de ejemplos internos almacenados en el cerebro, los cuales representan un grado de intensidad al que se le asigna un valor numérico (Feng *et al.*, 2015).

Para este modelo se exige un protocolo monádico, en el cual el juicio para un estímulo no se vea afectado por otros estímulos, es decir que no exista efecto de contexto. Lo anterior se logra al evaluar las muestras en días diferentes, no obstante, es poco práctico para la industria. Por lo cual si se realiza en un mismo día se recomienda eliminar el acceso a las respuestas una vez calificado (Feng *et al.*, 2015).

Este es el protocolo manejado al utilizar la escala hedónica de 9 puntos, aunque es más adecuado para instrumentos de laboratorio.

Al utilizar escalas hedónicas con frases se prefiere este modelo, debido a que se ajusta a la estrategia cognitiva que utiliza el cerebro en estas evaluaciones; la cual consiste en comparar el producto con controles internos almacenados previamente en el cerebro debido a la experiencia (Wichchukit & O'Mahony, 2015).

3.5.2. Modelo relativo de Mellers: Supone que el valor que se le da a un producto depende de los puntajes dados a los demás productos (Feng *et al.*, 2015). En este modelo se evidencian variaciones de la evaluación debido a las diferencias en el uso de la escala.

Este modelo es preferido cuando se hace uso de escalas puramente numéricas, como las escalas hedónicas numéricas, debido a que no existe efecto por el uso de la escala, sino que depende de las puntuaciones dadas a las demás muestras (Wichchukit & O'Mahony, 2015).

Así mismo, la utilización de un modelo relativo, cuando se hace uso de escalas numéricas, se fundamenta en el sesgo de igualdad de estímulo de Poulton; el cual se refiere al hecho de que los estímulos que son relativamente similares y los que son relativamente diferentes tienden a estar espaciados de la misma manera en toda la longitud de la escala (Wichchukit & O'Mahony, 2015).

Por otra parte, se presenta evidencia que demuestra que con esta estrategia cognitiva se evitan calificaciones erróneas por olvido al permitir al panelista la reevaluación de la muestras (Feng *et al.*, 2015).

### **3.6. Traducciones de la escala hedónica de 9 puntos al español**

La escala hedónica de 9 puntos se traduce al español haciendo uso de diferentes frases o palabras de acuerdo al país donde se utilice. En una investigación realizada en Argentina se construyeron tres escalas diferentes con el objeto de analizar el entendimiento de las traducciones al español de la escala hedónica de 9-puntos presentado por los consumidores argentinos (Curia *et al.*, 2001). En esta investigación se le presentó a cada panelista tres traducciones, una a una, con el fin de que el panelista acomodara las frases según su entendimiento con la posibilidad de invertir las frases; y de esta manera determinar cuál de las traducciones generó menor inversión de frases de acuerdo a la traducción establecida, es decir, cuál de las traducciones no presentaba mayor entendimiento para los consumidores argentinos.

Estudios similares se han realizado en México, donde se han construido traducciones de las frases de la escala hedónica de 9 puntos (Rosas *et al.*, 2005; Feng *et al.*, 2015). Las traducciones al español de la escala hedónica de 9 puntos utilizadas con consumidores argentinos y mexicanos se presentan en el Cuadro II.

**Cuadro II.** Categorías de la escala hedónica de 9 puntos en idioma inglés y cuatro traducciones al idioma español.

Inglés	Español			
	1*	2**	3**	4***
Like extremely	Gusta muchísimo	Me gusta extremadamente	Me gusta muchísimo	Gusta muchísimo
Like very much	Gusta mucho	Me gusta mucho	Me gusta mucho	Gusta mucho
Like moderately	Gusta moderadamente	Me gusta moderadamente	Me gusta bastante	Gusta moderadamente
Like slightly	Gusta poco	Me gusta apenas	Me gusta un poco	Gusta poco
Neither like or dislike	Ni gusta ni disgusta	Ni me disgusta ni me gusta	Ni me disgusta ni me gusta	Ni gusta ni disgusta
Dislike slightly	Disgusta poco	Me disgusta apenas	Me disgusta un poco	Disgusta poco
Dislike moderately	Disgusta moderadamente	Me disgusta moderadamente	Me disgusta bastante	Disgusta moderadamente
Dislike very much	Disgusta mucho	Me disgusta mucho	Me disgusta mucho	Disgusta mucho
Dislike extremely	Disgusta muchísimo	Me disgusta extremadamente	Me disgusta muchísimo	Disgusta muchísimo

\* (Rosas *et al.* 2005) \*\* (Curia *et al.*, 2001) \*\*\* (Feng *et al.*, 2015)

### **3.7. Limitaciones de la escala hedónica de 9 puntos**

#### **3.7.1. Uso de la escala, significado y orden de las frases de las categorías**

En estudios realizados con consumidores de diferentes culturas se ha evidenciado que se utiliza la escala hedónica de 9 puntos de forma diferente; por ejemplo, los consumidores asiáticos reducen el número de categorías en comparación a las que utilizan los estadounidenses. Lo anterior impide la posibilidad de comparar los resultados obtenidos por las diferentes poblaciones para un mismo producto (Feng & O'Mahony, 2017).

La utilización de la escala con traducción directa de la escala manejada en Estados Unidos varía en las diferentes culturas y genera ambigüedad en el significado de las frases (O'Mahony *et al.*, 2004). Curia *et al.* (2001) evidenciaron, en un estudio realizado en Argentina, que un 30 % de los analistas utilizaron la escala con distinto ordenamiento al presentárselas con traducción directa.

Yao *et al.* (2003) demostraron que la traducción de la escala a diferentes idiomas da resultados equivocados, esto al descubrir que para los coreanos la traducción de "extremely" a su idioma es más amplio en agrado que su significado en inglés.

#### **3.7.2. Espaciamiento psicológico de las frases, utilización de números y palabras**

Al utilizar la escala hedónica de 9 puntos se asume una diferencia de una unidad numérica entre las categorías consecutivas, lo cual hace que se incurra en sesgo debido a que al utilizar la escala únicamente con números se obtienen resultados diferentes que al emplear palabras, lo que confirma que los números utilizados en la escala hedónica de 9 puntos verbal no son intercambiables con los utilizados en estas mismas escalas con carácter numérico (Nicolas *et al.*, 2010; Feng *et al.*, 2015; Wichchukit & O'Mahony, 2015; Xia *et al.*, 2015), debido a las diferencias entre las distancias psicológicas de las frases al usar números o palabras (Lim, 2011). Esto se debe a que la estrategia cognitiva que utiliza el analista es diferente en cada caso (Villanueva *et al.*, 2005; Wichchukit & O'Mahony, 2015).

Debido a lo mencionado anteriormente, la escala hedónica de 9 puntos permite determinar cuáles muestras son más aceptadas, mas no permite

cuantificar la diferencia en agrado. Lo anterior se explica debido a que un puntaje de 8 al utilizar frases no es el doble de un puntaje de 4, lo que si sucede en escalas que utilizan números (Pimentel *et al.*, 2016). Por consiguiente, las categorías de la escala hedónica de 9 puntos deben presentar frases claras y graduales para cada uno de los panelistas que va a emplear el instrumento en la evaluación de los productos (González *et al.*, 2014).

### 3.7.3. Número limitado de categorías

La principal razón por la cual la escala hedónica de 9 puntos es la más usada en la industria alimentaria en pruebas de aceptación es que en comparación con otras escalas posee un número limitado de categorías, facilitando su uso en diferentes segmentos de la población. De igual manera se ha demostrado que en términos de poder de discriminación de la muestras es tan sensible como otras escalas (Pimentel *et al.*, 2016).

Por otro lado, Lim (2011) y Pimentel *et al.* (2016) concuerdan en que el número limitado de categorías en dicha escala genera dificultades para que el panelista exprese sus percepciones sensoriales, coincidiendo con Villanueva *et al.* (2005) y Simone & Pangborn (1957) que reportaron que en estudios realizados los consumidores manifestaron frustración al no poder expresar que, aunque dos productos estén en una misma categoría, tienen preferencia por uno de éstos.

Lo anterior concuerda con la afirmación de que al utilizar un tamaño de muestra mayor a las categorías existentes en la escala se genera sesgo en los resultados obtenidos dado que el panelista cuenta con menos categorías de las que pudiera necesitar (Angulo & O'Mahony, 2009). Contrario a esto, al presentar menor número de productos que categorías el panelista logra ubicar el producto en el sitio que considere de acuerdo a su agrado, y podría clasificar uno o más productos en una misma categoría si quisiera indicar que posee el mismo nivel de agrado por los productos. Por el contrario si son más muestras que categorías el panelista se ve obligado a colocar más de una muestra en una misma categoría aunque no representen el mismo nivel de agrado (Wichchukit & O'Mahony, 2015).

Al utilizar más muestras que categorías con una escala verbal de 9 puntos, que produce datos ordinales, se presenta el problema de agrupar dos muestras en una misma categoría que no tienen el mismo nivel de agrado. Para esto se

asume que se debe convertir el instrumento a una escala puramente numérica. Sin embargo, tener que clasificar una cantidad grande de productos genera problemas en cuanto al olvido del estímulo y se requeriría demasiada re-degustación. Como solución a los inconvenientes se recomienda clasificar los estímulos dentro de cada categoría, donde se requiere menor memoria; y posteriormente, para el análisis estadístico, se unen las clasificaciones y se hace uso de la estadística no paramétrica.

#### 3.7.4. Efecto de tendencia central

Se ha evidenciado que al utilizar esta escala se produce un efecto de tendencia central o también conocido como efecto final de la escala, el cual se define como la mayor dificultad que posee el panelista para pasar de la categoría 8 a la 9 y de la 2 a la 1, en comparación con las categorías centrales (Wichchukit & O'Mahony, 2015).

Este efecto consiste en evitar utilizar las categorías extremas y reducir la escala de 9 a 7 puntos, con lo cual la capacidad para detectar diferencias disminuye y se asocia a dos aspectos principales:

1. Número pequeño de categorías, por lo cual se prefiere utilizar las categorías centrales en comparación con las externas. Esto con la finalidad de reservar las categorías extremas (Lim, 2011) para que se encuentren disponibles si se prueba un producto que lo requiera y así brindar un puntaje adecuado a las muestras posteriores (Villanueva *et al.*, 2005; Pimentel *et al.*, 2016).
2. Espaciamiento desigual en categorías consecutivas (Wichchukit & O'Mahony, 2015).

De igual manera se asocia, en casos muy específicos, a un comportamiento cultural que hace que las personas no den más peso a unas respuestas sobre otras, lo cual fue evidenciado por Yeh *et al.* (1998) en un estudio realizado en consumidores asiáticos. Este comportamiento también ha sido demostrado por Yao *et al.* (2003) en un estudio con americanos, coreanos y japoneses. De igual manera se evidenció que la cultura asiática no permite dar calificaciones negativas a los productos y por consiguiente reducen la escala a valores de agrado y evitan las de desagrado.

### 3.7.5. Análisis estadístico

Al utilizar la escala hedónica de 9 puntos se obtienen datos categóricos, discretos y sin un punto cero, los cuales deben ser analizados mediante la estadística no paramétrica. Sin embargo, es muy común que se utilice estadística paramétrica, como el análisis de varianza, para obtener resultados a partir de los datos recolectados, lo cual es inapropiado matemáticamente debido a que se asumen aspectos como homogeneidad de las variaciones y normalidad de los datos (Pimentel *et al.*, 2016).

## 3.8. Análisis de datos

El análisis estadístico de los resultados obtenidos con la escala hedónica de 9 puntos se puede realizar mediante la estadística paramétrica o la no paramétrica, la cual es menos sensible, dependiendo del tratamiento que se le dé a los resultados obtenidos (Lim, 2011). Si el valor obtenido se clasifica como categórico y discreto se hace uso de la estadística no paramétrica, mientras que si se utilizan como datos continuos se hace uso de la estadística paramétrica (Jaeger & Cardello, 2009).

El análisis estadístico tradicional de los datos obtenidos mediante la escala hedónica de 9 puntos se fundamenta en la estadística paramétrica haciendo uso del análisis de varianza (ANDEVA) con la respectiva prueba de comparación de medias. Con este análisis se determina si existen diferencias significativas en el promedio de los puntajes a signados a las muestras (Ramírez, 2012).

### 3.8.1. Análisis de varianza (ANDEVA)

En el análisis tradicional de los resultados de la escala hedónica de 9 puntos se asigna a cada categoría un valor del 1 al 9 con el fin de utilizar la estadística paramétrica (Nicolas *et al.*, 2010). En este caso se le asigna el mayor valor al producto con mayor agrado y el menor al producto de menor agrado (Feng & O'Mahony, 2017).

Al atribuirle números a las categorías verbales se está convirtiendo la escala de ordinal a escala de intervalo, donde se asume que las diferencias entre las categorías consecutivas son iguales. Lo anterior se hace con el objetivo de obtener resultados numéricos y poder hacer una cuantificación de los resultados (Feng & O'Mahony, 2017).

En la ANDEVA, la varianza total se divide en varianza asignada a diferentes fuentes. La varianza de las medias entre muestras se compara con la varianza de dentro de la muestra (error experimental). Si las muestras no son diferentes, la varianza de las medias entre muestras será similar al error experimental (Watts *et al.*, 1989; Ramírez, 2012).

Los datos obtenidos mediante el uso de la escala hedónica de nueve puntos no cumple con los supuestos estadísticos de normalidad, homoscedasticidad y aditividad, requeridos en los modelos de análisis de varianza (Villanueva *et al.*, 2005; Lim & Fujimaru, 2010). Por consiguiente Wichchukit & O'Mahony (2015) afirman que el uso de la escala no es el causante de los problemas, sino el análisis estadístico que se realiza de los datos obtenidos a partir de ésta, al utilizarlos como si fueran generados por instrumentos calibrados y no por personas con diferentes capacidades cognitivas.

### 3.8.2. Índice R

El índice R se basa en la Teoría de Detección de Señales y es un estimador del área bajo la curva característica operacional de respuesta (ROC por sus siglas en inglés), lo cual corresponde a la probabilidad de que se elija un producto sobre el otro (Bi & O'Mahony, 2007). Este índice fue desarrollado por John Brown (1974) y se basa en la estadística no paramétrica. Ha sido utilizado para múltiples aplicaciones en análisis sensorial, entre las que se encuentran medidas de umbral (Robinson *et al.*, 2004), pruebas de diferencia (Lee & Van Hout, 2009), medidas de preferencia, medidas de aceptabilidad e intención de compra y conceptos abstractos del consumidor (Lee & O'Mahony, 2005). Lo anterior refleja la principal ventaja del índice R, la cual se basa en que es muy general y puede ser utilizado tanto en evaluaciones dirigidas a la obtención de calificaciones (rating) como las dirigidas a la generación de clasificaciones (ranking) (O'Mahony *et al.* 1980).

La teoría de detección de señales, o modelo Thurstoniano, se basa en la comprensión de que la percepción de un alimento es variable, lo cual se resalta como una ventaja de la utilización de este índice ante otras pruebas estadísticas. Así mismo, Lee & Van Hout (2009) establecen que el análisis del índice R tiene la ventaja de la simplicidad computacional y la flexibilidad en su capacidad de

aplicarse a una amplia variedad de protocolos de prueba aplicados en ciencia de alimentos.

El uso principal que se le dio a este estimador fue en las pruebas de diferencia, y fue para esto que se desarrolló la teoría de detección de señales. A través del índice R se obtiene la medida de la señal sensorial requerida para detectar diferencias entre muestras, esto mediante la probabilidad de identificar correctamente la muestra (señal) del control (ruido) independientemente del sesgo en la respuesta (Cliff *et al.*, 1997).

Posteriormente se realizaron modificaciones a la teoría de detección de señales con la finalidad de ser utilizada en evaluaciones hedónicas (Cliff *et al.*, 1997; Lee & Van Hout, 2009), para lo cual se recomendó inicialmente un tamaño base de al menos 100 panelistas para el panel de consumo típico en investigaciones de mercado (Basker, 1998). En cuanto a las pruebas hedónicas el índice R se basa en la probabilidad de elegir un alimento (señal) cuando se presenta en una prueba de preferencia contra otro producto (Lee & Van Hout, 2009).

En la obtención del índice R para pruebas hedónicas de consumidores se toman dos productos y se analiza la probabilidad de que uno de estos sea preferido en comparación a otro, para tal efecto se toma el producto líder como la señal y un producto de menor agrado como el ruido y se obtiene el índice para la señal. Posteriormente se puede obtener otro valor de índice R tomando el ruido anterior como señal y otro producto de menor agrado como el ruido, y así sucesivamente (Park *et al.*, 2007).

Para análisis de clasificación (ranking) se han establecido dos métodos diferentes de obtención del índice R, el  $R_{JB}$  y el  $R_{MAT}$ . El primero de estos lleva las siglas del nombre de su desarrollador (John Brown), se obtiene al calcular el porcentaje de veces que un producto antecede a otro en agrado y no requiere la construcción de matrices ni cálculos adicionales. Así mismo el  $R_{MAT}$  se calcula mediante la utilización de una matriz que posee el número de muestras que presentaron cada clasificación, lo anterior se construye para las dos muestras a analizar y haciendo uso de fórmulas matemáticas se obtiene el porcentaje.

No se ha demostrado que un método de obtención sea superior al otro en cuanto a su capacidad para detectar diferencias; sin embargo, se espera que el índice  $R_{JB}$  sea más discriminatorio debido al cálculo de cada método (Wichchukit

& O'Mahony, 2015). El  $R_{MAT}$  toma en cuenta la separación entre los productos a comparar y la posible existencia de otros estímulos entre los analizados, mientras que el  $R_{JB}$  únicamente toma en cuenta el orden de agrado de dos productos, es decir, el  $R_{JB}$  está libre del efecto de contexto y el  $R_{MAT}$  no. Sin embargo este último es una mejor estimación del área bajo la curva característica operacional de respuesta (ROC por sus siglas en inglés) (O'Mahony, 2017).

La utilización del índice R en pruebas de consumidor con escalas hedónicas para aspectos interculturales fue analizado por Feng & O'Mahony (2017) y establecieron que, tanto para estadounidenses como para chinos, se obtuvieron los mismos resultados, lo cual resalta la ventaja de este método de análisis de tener rango de puntuaciones iguales para ambos consumidores.

Por otra parte, Bi (2006) destaca algunas ventajas del uso de este análisis, entre las cuales se encuentran que el índice R es uno de los estadísticos no paramétricos más potentes y se prevé que es únicamente un 5% menos potente que la prueba t cuando se cumple la normalidad; sin embargo, cuando no se tiene la certeza de este aspecto es más potente el índice R.

De igual manera, el índice R es una medida basada en el comportamiento humano en lugar de estimaciones numéricas imperfectas obtenidas con una escala que funciona para mediciones ordinales en lugar de intervalo; y las calificaciones y clasificaciones obtenidas con pruebas al consumidor son consideradas medidas ordinales (Bi, 2006).

Así mismo, se destaca que este índice es una medida más directa, intuitiva e interpretable debido a que proporciona una estimación de la probabilidad de que dos productos sean distinguibles. De igual manera, este índice no se ve afectado por criterios de decisión o el número de categorías en los datos (Bi, 2006).

Otras ventajas presentes en la utilización de este índice es que al ser no paramétrico se evita la necesidad de cumplir supuestos de normalidad que si son requeridos en la estadística paramétrica y no solamente indica que dos productos son diferentes sino que da un tamaño de diferencia (Varela *et al.*, 2014). Aunado a esto se destaca que en la utilización del índice R se tiene claro los límites y distancias entre los rangos (Feng & O'Mahony, 2017).

Entre las limitaciones de la utilización del índice R se encuentra que es propenso a variar con el método experimental, por lo que la metodología a utilizar

en las pruebas, tanto de discriminación como hedónicas, debe ser muy controlada (Lee & Van Hout, 2009).

### **3.8.3. Interpretación del análisis de resultados de la escala hedónica de 9 puntos**

Al realizar la interpretación de los resultados obtenidos con la escala hedónica de 9 puntos se incurre en errores debido al uso idiosincrático de la escala. Lo anterior hace referencia a que los jueces utilizan la escala hedónica de diferentes formas, por lo que no se puede comparar a los jueces en cuanto a las calificaciones debido a que éstas no son absolutas, sino relativas. Si un juez utiliza las categorías en la posición 8 y 9 para dos productos no quiere decir que sea una distancia diferente que la expresada por otro juez que coloca los productos en las posiciones 4 y 5 (Wichchukit & O'Mahony, 2015).

De igual manera, en la interpretación de los resultados por parte de la industria alimentaria se hace uso del análisis denominado en inglés "top two box", lo cual hace referencia a que un producto debe estar ubicado en las dos posiciones de mejor calificación, sino será rechazado (O'Mahony, 2017). Al utilizar la escala hedónica numérica de 9 puntos se requiere un valor de 8 o 9 para decir que un producto no será rechazado, mientras que en la escala verbal se requiere que el producto esté en las categorías 'like extremely' o 'like very much' para que sea aceptado.

No obstante, al analizar la escala verbal se convierten los resultados en números y, por tanto, se asume que sólo se deben aceptar los productos con puntuación mayor a 7, lo cual no es correcto y puede no ser significativo, por consiguiente se incurre en un error al momento de analizar los resultados (Brody & Lord, 2007).

## **4. MATERIALES Y MÉTODOS**

### **4.1. Localización**

Las pruebas se llevaron a cabo en el Laboratorio de Análisis Sensorial de la Escuela de Tecnología de Alimentos de la Universidad de Costa Rica, Sede Rodrigo Facio, Montes de Oca, San José.

### **4.2. Escogencia de productos: prueba preliminar**

Se probaron bizcochos y gomitas de una amplia variedad de marcas. Un grupo de cuatro personas probó 18 gomitas de sabores frutales y 5 bizcochos de maíz, para determinar el producto a utilizar. Se buscó el producto que generara sensaciones extremas con la finalidad de lograr con mayor seguridad impresiones de agrado y rechazo para que se utilizara la escala en toda su amplitud.

Se seleccionó las gomitas que presentaron mayor variedad y diferencias en cuanto a las características generales del producto; y se descartaron los bizcochos porque no presentaron muchas variaciones en el sabor y textura.

Las gomitas se clasificaron en cinco diferentes grupos, como se muestra en el Cuadro III, para facilidad del manejo de la información.

Las gomitas dentro de una clasificación eran similares en sabor y se diferenciaban de las otras clasificaciones. Se escogió un producto de cada categoría (5 muestras) para utilizarlos en el análisis con menos muestras que categorías, lo cual se muestra en el Cuadro V.

**Cuadro III.** Muestras de gomitas analizadas en la escogencia del producto a utilizar; con su respectiva marca, clasificación y observaciones.

<b>Clasificación</b>	<b>Gomita</b>	<b>Marca</b>	<b>Observaciones</b>
1	Gummy Bears	VIDAL	Dura, sólo ácida y sin sabor
	Banana Banana	GUANDY	Feo
	Teeth	VIDAL	Sólo azúcar, dura, parece marshmellow
	Aros	Trululu	Forma y color
2	Watermelon slices	VIDAL	Ácido
	Jelly Sharks	VIDAL	Tiburones de color
	Ositos Scout Bears	Tropical	Ositos
3	Culebritas Sour Snakes	GUANDY	No tan mal
	Strawberry & cream Drops	VIDAL	Suaves de color rosado y blanco
	Pingüinitos	Ricolino	Pingüinos negro con blanco
4	Gomilocas Dientes	Ricolino	Poco ácido, gustó
	Gomitas	Diana	Dulces
	Dulcigomas	Ricolino	Más Dulces
	Gummy Bears	Black forest	Similar a marca VIDAL, sin azúcar
5	Fresitas	Trululu	Rosado fuerte
	DINOSS	Trululu	De las más gustadas
	Panditas	Ricolino	Similar a Gummy Bears
	Goma Pino	Diana	La que más gustó

De igual manera, se eliminaron cinco muestras las cuales presentaron similitudes en forma visual con otras, de manera que se retuvieron 13 gomitas para utilizarlas en el análisis con mayor número de muestras que categorías; los cuales se encuentran en el Cuadro VI.

### 4.3. Traducción de términos de la escala

Con la ayuda de una especialista en traducción se plantearon traducciones al español de cada una de las categorías de la escala hedónica de nueve puntos, en las cuales se buscó que estas frases se ajustaran al lenguaje más común del costarricense.

A un grupo de 50 personas se les dio las traducciones de cada una de las categorías de la escala hedónica de 9 puntos y se les pidió que a partir de ellas construyeran una, dos o tres escalas donde el valor central fuera un punto intermedio entre el agrado y el desagrado y aumentara hacia los extremos. Lo anterior con el fin de seleccionar la escala que se presentara mayor cantidad de veces (Anexo 9.5). El Cuadro IV muestra las traducciones de cada una de las categorías presentadas.

**Cuadro IV.** Categorías de escala hedónica de 9 puntos en inglés con sus respectivas traducciones presentadas durante la búsqueda de frases para la elaboración de la escala en el idioma español.

<b>Categoría en inglés</b>	<b>Traducciones</b>
Like extremely	Me gusta demasiado
	Me gusta extremadamente
	Me gusta muchísimo
	Me gusta excesivamente
	Me encanta
Like very much	Me fascina
	Me gusta mucho
	Me gusta bastante
Like moderately	Me gusta tanto
	Me gusta moderadamente
	Me gusta razonablemente
	Me gusta mesuradamente
	Me gusta regularmente
	Me gusta pero no tanto

**Continuación cuadro IV.** Categorías de escala hedónica de 9 puntos en inglés con sus respectivas traducciones presentadas durante la búsqueda de frases para la elaboración de la escala en el idioma español.

<b>Categoría en inglés</b>	<b>Traducciones</b>
Like slightly	Me gusta apenas
	Me gusta escasamente
	Me gusta apenas un poco
	Me gusta raramente
	Me gusta un poco
Neither like or dislike	No me gusta ni me disgusta
	Me es indiferente
Dislike slightly	Me disgusta un poco
	Me disgusta apenas un poco
	Me disgusta raramente
	Me disgusta escasamente
	Me disgusta apenas
Dislike moderately	Me disgusta moderadamente
	Me disgusta razonablemente
	Me disgusta mesuradamente
	Me disgusta regularmente
	Me disgusta pero no tanto
Dislike very much	Me disgusta mucho
	Me disgusta bastante
	Me disgusta tanto
Dislike extremely	Me disgusta extremadamente
	Me disgusta demasiado
	Me disgusta muchísimo
	Me disgusta excesivamente

Se permitió a las personas agregar otras frases adicionales para las categorías o utilizar las frases en un orden distinto, es decir, utilizar una frase en español para representar una categoría en inglés. Se contó el número de personas que generó las diferentes escala y así se determinó cuál es el orden que mayor número de personas escogió (Yeh *et al.*, 1998; Hough *et al.*, 2006).

#### 4.4. Evaluación del agrado de los productos usando menos muestras que categorías

Mediante la utilización de la escala hedónica de nueve puntos obtenida en la sección 4.3 se procedió a evaluar cinco muestras.

##### 4.4.1. Escala

Se utilizaron las nueve frases de la escala que se escogieron con base en la sección 4.3. La presentación de la escala fue física, en forma de cajones que representaban las categorías para que las muestras se colocaran dentro de la celda correspondiente y no se escribió ni se usó la computadora. En la Figura 1 se muestra la escala utilizada en la evaluación.



**Figura 1.** Escala hedónica de 9 puntos utilizada en las evaluaciones sensoriales de agrado de gomitas.

La escala fue construida con las siguientes características: 70 cm de largo y 40 cm de ancho, un tamaño de las columnas uniforme (7,78 cm), tipo de letra Agency FB y tamaño de 11,59 mm, de manera que físicamente hubo suficiente espacio para colocar las muestras en cualquiera de las categorías que los consumidores seleccionaran.

##### 4.4.2. Productos

Se utilizaron las siguientes cinco muestras que se presentan en el Cuadro V.

**Cuadro V.** Muestras de gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos.

Producto	Marca
Goma Pino	DIANA
Watermelon slices	VIDAL
Strawberry & cream drops	VIDAL
Dulcigomas	Ricolino
Teeth	VIDAL

### 4.4.3. Evaluación

#### 4.4.3.1. Consumidores

Se incorporaron 200 panelistas, 65 hombres y 135 mujeres, de edades entre 18 y 35 años; que son consumidores regulares de gomitas y evaluaron los cinco productos en una sesión. Al juez (consumidor) se le acompañó de manera que entendiera el uso de la escala y no tuviera que escribir ni marcar en la escala sino que esa fue tarea del evaluador; el fin es que no se distrajera escribiendo.

Cada panelista probó la muestra y colocó la gomita en el espacio correspondiente a la categoría que representaba el nivel de agrado del producto, con la posibilidad de colocar dos productos en una misma categoría y dejar categorías vacías. Posteriormente, si se había colocado más de un producto en una categoría, se le solicitaba al panelista que ordenara estos productos de mayor a menor en orden de preferencia, con la posibilidad de presentar empates en cuanto al agrado de los productos.

### **4.5. Evaluación del agrado de los productos usando más muestras que categorías**

Mediante la utilización de la escala hedónica de nueve puntos obtenida (sección 4.3) se procedió a evaluar trece muestras, haciendo uso de los siguientes componentes.

#### 4.5.1. Escala

Se utilizaron las nueve frases de la escala que se escogieron en la sección 4.3 como se muestra en la Figura 1. Al haber más muestras que categorías se permitió asignar varias de las muestras analizadas a una misma categoría y dejar categorías vacías. Posteriormente, a esas muestras de una misma categoría se les pidió que las ordenaran de mayor a menor agrado.

#### 4.5.2. Productos

Se analizaron visualmente 13 productos con pequeñas variaciones, debido a que es un número alto de muestras con sabores fuertes, y podría darse fatiga, adaptación o acarreamiento si se probaban, por lo que no se llevó a cabo la evaluación por medio del sabor. Ya que el estudio enfatizaba el uso de la escala y el análisis de datos, es decir no interesa el producto *per se*, se hizo un análisis visual. Los productos que se evaluaron se encuentran en el siguiente Cuadro.

**Cuadro VI.** Muestras de gomitas utilizadas en la evaluación de más muestras que categorías mediante la escala hedónica de 9 puntos.

<b>Producto</b>	<b>Marca</b>
Aros	Trululu
Gomilocas Dientes	Ricolino
Teeth	VIDAL
Gomitas	DIANA
Culebritas Sour Snakes	GUANDY
Banana Banana	GUANDY
Gomilocas Pingüinos	Ricolino
Watermelon slices	VIDAL
SHARKS	VIDAL
DINOSS	Trululu
Fresitas	Trululu
Goma Pino	DIANA
Ositos Scout Bears	Tropical

Como se indicó en la sección 4.2, estas muestras tenían formas, textura aparente y colores diferentes, de manera que se pudo garantizar que algunas muestras se rechazaran y otras fueran muy gustadas por su apariencia.

#### 4.5.3. Evaluación

##### 4.5.3.1. Consumidores

Doscientos consumidores, 69 hombres y 131 mujeres con edades entre 17 y 59 años, evaluaron visualmente los trece productos en una sesión, en cuanto a agrado general. Al juez (consumidor) se le acompañó de manera que entendió el uso de la escala y no tuvo que escribir ni marcar en la escala sino que esa fue tarea del evaluador; el fin era que no se distrajera el participante escribiendo.

Cuando terminó la evaluación anterior se le pidió al participante que ordenara las muestras ubicadas en una misma categoría de mayor a menor de acuerdo al nivel de agrado, esto se hizo con cada categoría donde hubo varias muestras. Al final se obtuvieron trece categorías de una escala de ordenamiento; ya no era una escala de categorías cerrada.

## 4.6. Análisis

### 4.6.1. Análisis de datos tradicional

Las categorías se transformaron a números siguiendo las indicaciones que se presentan en el Cuadro VII.

**Cuadro VII.** Valor numérico asignado a cada una de las categorías de la escala hedónica de nueve (Wichchukit & O'Mahony, 2015).

<b>Categoría en inglés</b>	<b>Valor asignado</b>
Like extremely	9
Like very much	8
Like moderately	7
Like slightly	6
Neither like or dislike	5
Dislike slightly	4
Dislike moderately	3
Dislike very much	2
Dislike extremely	1

Los datos de la escala hedónica, con 5 productos y con 13 productos, se utilizaron de acuerdo con el valor numérico de la categoría donde se ubicó y se realizó un análisis de varianza (ANDEVA) con  $\alpha$  de 0,05 y una comparación de medias (Fisher LSD) cuando se encontró diferencias significativas y con  $\alpha$  de 0,05, siendo éste el análisis tradicional de este tipo de datos.

### 4.6.2. Análisis mediante índice R de ordenamiento

La obtención del índice R de ordenamiento se realizó mediante dos métodos,  $R_{MAT}$  y  $R_{JB}$ .

#### 4.6.2.1. R<sub>JB</sub>

##### 4.6.2.1.1. Menos muestras que categorías (5 gomitas)

1. Codificación de muestras: Se codificaron las muestras utilizadas, como se muestra en el Cuadro VIII

**Cuadro VIII.** Muestras de gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos y su código respectivo.

Producto	Código
Goma Pino	P
Watermelon slices	W
Strawberry & cream drops	S
Dulcigomas	D
Teeth	T

2. Se realizó un ordenamiento de los productos evaluados, dándole el valor de 1 al producto con mayor agrado y un valor de 5 al producto con menor agrado. En caso de presentarse empates entre productos se promedió las clasificaciones correspondientes para éstos y se les asignó el mismo valor a ambos. Se repite para cada uno de los 200 consumidores y se obtiene una tabla de 202 filas y 10 columnas, como se muestra en el Cuadro IX.

**Cuadro IX.** Clasificación de las categorías del ordenamiento de las gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos por cada juez, donde 1º fue la de mayor agrado y los valores mayores menos agradables.

Juez	Clasificación								
	1	1,5	2	2,5	3	3,5	4	4,5	5
1									

3. Se contabilizó la cantidad de veces (número de consumidores) que un producto presentó mayor agrado (número menor en la escala) con respecto a otro (número mayor). Esto se realizó para cada muestra con respecto a las demás; es decir por pares de muestras.

4. Se obtuvo el índice R de ordenamiento de un producto sobre otro, siguiendo la ecuación 1 y se obtuvieron 9 índices  $R_{JB}$  en total para el análisis de las 5 muestras.

$$P_1 \rightarrow P_2 = \frac{X}{N} * 100 \quad (1)$$

Donde:

$P_1$ : Producto 1 (mayor agrado)

$P_2$ : Producto 2 (menor agrado)

X: Cantidad de veces que un  $P_1$  presenta mayor agrado con respecto a  $P_2$

N: Número total de jueces

5. Se comparó el resultado para cada par de muestras con el valor de las tablas de Bi *et al.* (2007) con un  $\alpha$  de 0,05 y dos colas para 200 consumidores (55,62 %). La muestra de cálculo se presenta en el Cuadro XXVIII de Anexos.

#### 4.6.2.1.2. Más muestras que categorías (13 gomitas)

1. Codificación de muestras: Se codificaron las muestras utilizadas, las cuales se presentan en el Cuadro X.

**Cuadro X.** Muestras de gomitas utilizadas en la evaluación de más muestras que categorías mediante la escala hedónica de 9 puntos y su código respectivo.

Producto	Código
Aros	A
Gomilocas Dientes	R
Teeth	T
Gomitas	G
Culebritas Sour Snakes	C
Banana	B

**Continuación Cuadro X.** Muestras de gomitas utilizadas en la evaluación de más muestras que categorías mediante la escala hedónica de 9 puntos y su respectivo código.

Producto	Código
Gomilocas Pingüinos	Y
Watermelon slices	W
SHARKS	V
DINOSS	Z
Fresitas	F
Goma Pino	P
Ositos Scout Bears	O

- Se realizó un ordenamiento de los productos evaluados, dándole el valor de 1 al producto con mayor agrado y un valor de 13 al producto con menor agrado. En caso de presentarse empates entre productos se promedió las clasificaciones correspondientes para éstos y se les asigna el mismo valor a ambos. El Cuadro XI muestra el formato de la tabla obtenida a partir del ordenamiento, donde se coloca cada letra correspondiente al código del producto de acuerdo a la posición en la escala.

**Cuadro XI.** Clasificación de las categorías del ordenamiento de las gomitas utilizadas en la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos por cada juez.

Juez	Clasificación																								
	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10	10,5	11	11,5	12	12,5	13
1																									

- Se contabilizó la cantidad de veces (número de consumidores) que un producto presentó mayor agrado (número menor) con respecto a otro (número mayor). Esto se realizó para cada muestra con respecto a cada una de las demás (par de muestras).
- Se obtuvo el porcentaje de preferencia de un producto sobre otro, siguiendo la ecuación 1 y se obtuvieron 78 índices  $R_{JB}$ , para la evaluación de las 13 muestras.

- Se comparó el resultado para cada par de muestras con el valor de las tablas de Bi y O'Mahony (2007) con un  $\alpha$  de 0,05 y dos colas para 200 consumidores (55,62 %).

#### 4.6.2.2. $R_{MAT}$

##### 4.6.2.2.1. Menos muestras que categorías (5 gomitas)

- Codificación de las muestras: Se codificaron las muestras como se aprecia en el Cuadro VIII.
- Ordenamiento por preferencia: Para cada consumidor se asignó a cada código de la muestra, del Cuadro VIII, el valor correspondiente en el orden de agrado, los cuales son números entre 1 y 5, donde 1 representa el producto de mayor agrado y 5 el de menor agrado. En caso de presentarse empates entre productos se promedió las clasificaciones correspondientes para estos y se les asignó el mismo valor a ambos.

El Cuadro XII muestra el formato de la tabla obtenida para las muestras analizadas.

**Cuadro XII.** Clasificación de los cinco productos evaluados mediante la escala hedónica de 9 puntos por cada juez.

Juez	Productos				
	P	W	S	D	T
1	1 <sup>o</sup>	4 <sup>o</sup>	5 <sup>o</sup>	2 <sup>o</sup>	3 <sup>o</sup>

- Se calculó la cantidad de veces que un producto se encontraba en cada posición, y se construyó la matriz del Cuadro XIII (ver muestra en Cuadro XXVI de Anexos).

**Cuadro XIII.** Posicionamiento (ordenamiento) de dos de los productos evaluados mediante la escala hedónica de 9 puntos.

Producto	Calificación								
	1	1,5	2	2,5	3	3,5	4	4,5	5
P	P1	P2	P3	P4	P5	P6	P7	P8	P9
W	W1	W2	W3	W4	W5	W6	W7	W8	W9

En los espacios (P1-P9, W1-W9) se anota el número de veces (cuántos consumidores) que el producto P o W se ubicó en la posición correspondiente.

4. Comparación entre productos: se obtuvo el porcentaje de jueces que prefieren un producto en comparación con otro, para ello se comparó la primera columna del producto 1 contra las columnas siguientes del producto 2 y así sucesivamente para cada posición.

$$\begin{aligned}
 P_W = & P1 * (W2 + \dots + W9) + P2 * (W3 + \dots + W9) + P3 * (W4 + \dots + W9) + P4 * \\
 & (W5 + \dots + W9) + P5 * (W6 + \dots + W9) + P6 * (W7 + \dots + W9) + P7 * \\
 & (W8 + \dots + W9) + P8 * W9 + \frac{1}{2} * (P1 * W1 + P2 * W2 + P3 * W3 + P4 * \\
 & W4 + P5 * W5 + P6 * W6 + P7 * W7 + P8 * W8 + P9 * W9) + \frac{1}{2} (P1 * W1 + \\
 & P2 * W2 + P3 * W3 + P4 * W4 + P5 * W5 + P6 * W6 + P7 * W7 + P8 * W8 + \\
 & P9 * W9)
 \end{aligned}
 \tag{2}$$

$$\%P_W = \frac{P_W}{N * N} * 100
 \tag{3}$$

Donde:

N: Número total de jueces

5. Se comparó el resultado para cada par de muestras con el valor de las tablas de Bi *et al.* (2007) con un  $\alpha$  de 0,05 y dos colas para 200 consumidores (55,62%). Una muestra de cálculo se encuentra en el Cuadro XXIX de Anexos.

#### 4.6.2.2.2. Más muestras que categorías (13 gomitas)

1. Codificación de las muestras: Se codificaron las muestras como se aprecia en el Cuadro IX.

2. Ordenamiento por preferencia: Para cada consumidor se asignó a cada código el valor correspondiente de ordenamiento, los cuales fueron números entre 1 y 13, donde 1 representaba el producto de mayor agrado y 13 el de menor agrado. En caso de presentarse empates entre productos se promedió las clasificaciones correspondientes para estos y se les asignó el mismo valor a ambos. Se elaboró la tabla que se aprecia en el Cuadro XIV.

**Cuadro XIV.** Clasificación por ordenamiento de los trece productos evaluados mediante la escala hedónica de 9 puntos por cada juez.

Juez	Producto												
	A	R	T	G	C	B	P	W	V	Z	F	Y	O
1	2º	7º	9º	1º	10º	11º	8º	3º	13º	4º	12º	5º	6º

3. Se calculó la cantidad de veces (cuántos consumidores) que un producto se encontraba en cada posición.

El Cuadro XV ejemplifica la matriz elaborada a partir de los resultados obtenidos para la realiza del análisis de datos.

**Cuadro XV.** Posicionamiento (ordenamiento) de dos de los productos evaluados mediante la escala hedónica de 9 puntos.

Producto	Clasificación																										
	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10	10,5	11	11,5	12	12,5	13		
A	A1																									A25	
D	D1																										D25

En los espacios (A1-A25, D1-D25) se anota el número de veces que el producto A o D se ubicó en la posición correspondiente.

4. Comparación entre productos: se obtuvo el porcentaje de jueces que prefieren un producto en comparación con otro.

$$\begin{aligned}
A_D = & A1 * (D2 + \dots + D25) + A2 * (D3 + \dots + D25) + A3 * (D4 + \dots + D25) + A4 * \\
& (D5 + \dots + D25) + A5 * (D6 + \dots + D25) + A6 * (D7 + \dots + D25) + A7 * \\
& (D8 + \dots + D25) + \dots + A24 * D25 + \frac{1}{2} * (A1 * D1 + A2 * D2 + A3 * D3 + A4 * \\
& D4 + A5 * D5 + A6 * D6 + A7 * D7 + A8 * D8 + \dots + A25 * D25) \quad (4)
\end{aligned}$$

$$\%A_D = \frac{A_D}{N*N} * 100 \quad (5)$$

5. Se comparó el resultado para cada par de muestras con el valor de las tablas de Bi *et al.* (2007) con un  $\alpha$  de 0,05 y dos colas para 200 consumidores (55,62%).

#### 4.6.3. Comparación de resultados

Se realizó la comparación de los resultados obtenidos mediante el índice R ( $R_{JB}$  y  $R_{MAT}$ ) para ordenamiento y el ANDEVA con  $\alpha$  de 0,05; de manera que se podía visualizar los cambios en conclusiones que se obtuvieron con cada análisis de datos (paramétrico comparado con no paramétrico).

#### 4.6.4. Comparación de los índices $R_{JB}$ y $R_{MAT}$ para dos tamaños de muestras

Se contabilizó el número de veces que la magnitud de un índice ( $R_{MAT}$  o  $R_{JB}$ ) superaba al otro, tomando en cuenta únicamente las parejas de comparaciones que detectaron diferencias por, al menos, uno de los dos índices R; es decir, se analizaron todas las diferencias significativas detectadas por los dos métodos. Luego se tomaron estas parejas de productos y se comparó con cuál índice R se obtenía mayor número de veces una diferencia significativa.

## 5. RESULTADOS Y DISCUSIÓN

### 5.1. Escala hedónica de 9 puntos traducida al español

Las frases de la escala hedónica de 9 puntos debieron establecerse para la población en estudio con el fin de evitar errores en las evaluaciones (Figura 2 y 3). El Cuadro XVI muestra la secuencia de la escala hedónica de 9 puntos en inglés, así como su traducción directa al español y la escala establecida por los consumidores para la evaluación de las muestras.

**Cuadro XVI.** Categorías de la escala hedónica de 9 puntos en inglés, traducción directa al español y escala obtenida para la evaluación de los productos.

Categorías en Inglés	Categorías en español	
	Traducción directa	Establecida por los consumidores
Like extremely	Me gusta extremadamente	Me gusta muchísimo
Like very much	Me gusta mucho	Me gusta mucho
Like moderately	Me gusta moderadamente	Me gusta moderadamente
Like slightly	Me gusta poco	Me gusta un poco
Neither like or dislike	No me gusta ni me disgusta	No me gusta ni me disgusta
Dislike slightly	Me disgusta poco	Me disgusta un poco
Dislike moderately	Me disgusta moderadamente	Me disgusta moderadamente
Dislike very much	Me disgusta mucho	Me disgusta mucho
Dislike extremely	Me disgusta extremadamente	Me disgusta muchísimo

La secuencia que el mayor número de panelistas utilizó en la construcción de la escala hedónica de 9 puntos en el estudio fue equilibrada en el disgusto y gusto, es decir, utilizaron las mismas frases para niveles opuestos de agrado. Así mismo, esta secuencia coincide con la seleccionada en dos investigaciones para ser utilizadas con consumidores mexicanos (Rosas *et al.*, 2005; Feng *et al.*, 2015). Lo anterior no implica que la escala seleccionada por los consumidores del estudio sea correcta o no, únicamente da una referencia de las frases que han sido empleadas en países de habla hispana y apoya a investigadores que

establecen que las traducciones varían en las diferentes culturas y genera ambigüedad en el significado de las frases (O'Mahony *et al.*, 2004).

La escala elaborada tiene similitudes con la escala utilizada por consumidores argentinos (Curia *et al.*, 2001); no obstante, presenta variaciones en cuatro categorías (Cuadro II) y con esto también se demuestra que las frases no tienen el mismo uso en diferentes países (Yao *et al.*, 2003). Por lo que, la construcción de una escala adaptada al lenguaje más común del costarricense permite una mejor utilización de dicha herramienta.

El Cuadro XVI muestra que la traducción de la escala hedónica de 9 puntos al español construida por los consumidores no coincidió con la escala hedónica de 9 puntos traducida de forma directa al idioma español, debido a las variaciones presentadas en los extremos de la escala que evidencian que no es común utilizar la palabra “extremadamente”, como si lo es usar “muchísimo”. Lo anterior evidenció que al utilizar la escala obtenida de la traducción directa, como se emplea con frecuencia (Castañeda, 2013), se está incurriendo en un error debido a que no utiliza el lenguaje que el consumidor emplea para ciertas categorías y, por consiguiente, la distancia entre las frases no es la misma (Nicolas *et al.*, 2010; Feng *et al.*, 2015; Wichchukit & O'Mahony, 2015; Xia *et al.*, 2015).

En este caso las personas no utilizaron frases dadas para una categoría en alguna de las otras categorías, aunque se les diera la opción. Por lo anterior, no se presentó inversión de las frases como sucedió en un estudio realizado con consumidores argentinos (Curia *et al.*, 2001), los cuales presentaron inversión en una o dos frases.

## **5.2. Análisis estadístico paramétrico y no paramétrico de la evaluación de más muestras que categorías**

### **5.2.1. Análisis tradicional (paramétrico)**

El Cuadro XVII muestra los resultados obtenidos con el análisis de varianza, el valor F y la probabilidad asociada, del agrado de 13 muestras de gomitas evaluadas visualmente utilizando la escala hedónica de 9 puntos.

**Cuadro XVII.** Análisis de varianza para determinar diferencias en el agrado de 13 muestras de gomitas mediante la escala hedónica de 9 puntos.

<b>F</b>	<b>Pr &gt; F</b>
62,717	< 0,0001

Se aprecia que el agrado generó un valor de probabilidad menor a 0,0001, es decir, hubo diferencia significativa en el agrado de al menos dos de las 13 muestras evaluadas visualmente mediante la escala hedónica de 9 puntos.

El Cuadro XVIII muestra la comparación de medias obtenidas de la evaluación de las 13 muestras mediante la utilización de la escala hedónica de 9 puntos, evidenciando las diferencias entre éstas.

**Cuadro XVIII.** Resultados de la comparación de medias de las calificaciones dadas por los consumidores a las trece muestras analizadas.

<b>Producto</b>	A	W	Z	V	F	Y	P	O	C	B	R	G	T
<b>Promedio</b>	7,21 <sup>a</sup>	7,16 <sup>a</sup>	7,06 <sup>a</sup>	6,91 <sup>a</sup>	6,19 <sup>b</sup>	6,16 <sup>b</sup>	6,09 <sup>b</sup>	5,93 <sup>b</sup>	5,85 <sup>bc</sup>	5,54 <sup>c</sup>	4,88 <sup>d</sup>	4,45 <sup>e</sup>	3,84 <sup>f</sup>

Letras diferentes indican diferencias significativas con  $p \leq 0,05$

En el Cuadro XVIII se muestra, de forma decreciente, el orden del agrado de los trece productos analizados visualmente por los consumidores; de izquierda a derecha (A-W-Z-V-F-Y-P-O-C-B-R-G-T); obtenidos mediante el análisis de los datos tradicional (ANDEVA). De igual manera se aprecia que los cuatro productos de mayor agrado (A-W-Z-V) no presentaron diferencias significativas entre ellos, mientras que los cuatro de menor agrado (B-R-G-T) eran diferentes significativamente entre ellos. Esto indicó que había gomitas agradables a la vista que no sobresalieron mientras que las desagradables eran fácilmente juzgadas y diferenciadas.

Con el análisis tradicional de los datos se presentó dificultad para detectar las diferencias en los productos de mayor agrado, se observa que el agrado de los primeros cuatro productos (A, W, Z, V), así como los cinco siguientes (F, Y, P, O, C) no fue, en cada caso, diferente estadísticamente.

El problema en la detección de las diferencias se asocia a varios aspectos, entre los cuales se encuentra principalmente que al presentarse mayor cantidad de muestras que de categorías se presenta sesgo en los resultados obtenidos

debido a que no existe espacio suficiente en la escala para discriminar trece productos (O'Mahony, 2017); es decir, el panelista cuenta con menos categorías de las que pudiera necesitar (Angulo & O'Mahony, 2009). Lo anterior se ha evidenciado en estudios previos, en los cuales los consumidores han manifestado frustración al no poder expresar que, aunque dos productos estén en una misma categoría, tienen preferencia por uno de éstos (Simone & Pangborn, 1957; Villanueva *et al.*, 2005).

Otro aspecto relevante es que el ANDEVA se basa en determinar si existen diferencias en el promedio del puntaje de la muestra (Ramírez, 2012), para lo cual se requiere atribuirle números a las categorías verbales y convertir la escala de ordinal a escala de intervalo (O'Mahony, 2017); donde se asume que las diferencias entre las categorías consecutivas son iguales, lo cual no es correcto al utilizar escalas hedónicas (Feng & O'Mahony, 2017).

El uso idiosincrático que se da de la escala hedónica de nueve puntos es otro factor que limita la visualización de las diferencias, por lo que no se puede comparar a los jueces en cuanto a las calificaciones debido a que éstas no son absolutas, sino relativas. Los promedios obtenidos no necesariamente reflejan la posición correspondiente en cuanto a agrado (Wichchukit & O'Mahony, 2015), debido a que un puntaje intermedio puede generarse a causa de que la mitad de las evaluaciones sean la máxima y la otra mitad la mínima, mientras que otro producto puede tener un valor central porque la totalidad de las evaluaciones le asignaron ese valor; y al realizar la comparación de promedios no se encuentran diferencias (Feng & O'Mahony, 2017).

Se aprecia que el valor máximo está cercano a 7 y el valor mínimo próximo a 4, lo cual con el análisis tradicional indica que en promedio el producto de mayor agrado se ubica entre "me gusta mucho" y "me gusta moderadamente", mientras que el producto con menor agrado se clasifica entre "me disgusta poco" y "me disgusta moderadamente". La ubicación de los productos en la escala no debe influir en la decisión de si un producto debe ser aprobado o rechazado. Si se tomara en cuenta el análisis denominado en inglés "top two box" se asumiría que la totalidad de los productos deben ser rechazados debido a que están en posiciones inferiores a las dos calificaciones de mayor agrado (Brody & Lord, 2007), no obstante, se debe tomar en cuenta el uso idiosincrático de la escala y su influencia en el promedio de los datos (O'Mahony, 2017).

De igual manera se observa que diez productos se encuentran en la zona de gusto mientras que los tres productos restantes se agrupan en el área de disgusto, por lo cual si se utiliza el análisis “top two box” se estaría catalogando la totalidad de los productos con agrado menor al que realmente presentaban y serían rechazados.

## 5.2.2. Índice R para ordenamiento

### 5.2.2.1. $R_{MAT}$

El Cuadro XIX muestra los valores del índice  $R_{MAT}$  donde se compara cada muestra de la primera fila contra cada una de la primera columna (matriz). El orden de agrado de los productos se define en la primera fila y la primera columna, y se obtiene un valor de índice R para determinar si existe diferencia entre un producto (primera fila) y los que se encuentran en una posición inferior de agrado (primera columna). Para productos que no presentan diferencias se tiene un valor de 50,00% y entre más cercano a 100% sea el valor mayor es la diferencia entre los productos analizados. Es por esto que los valores van aumentando, siendo mayores al comparar productos que se encuentran más distanciados de acuerdo al agrado.

**Cuadro XIX.** Índice  $R_{MAT}$  para ordenamiento obtenido para trece productos mediante la utilización de la escala hedónica de 9 puntos.

Primero \ Segundo	A	W	Z	V	F	P	Y	O	C	B	R	G
W	50,11											
Z	51,06	51,00										
V	<b>56,40</b>	<b>56,05</b>	55,20									
F	<b>69,13</b>	<b>67,37</b>	<b>64,46</b>	<b>62,37</b>								
P	<b>65,31</b>	<b>64,48</b>	<b>62,87</b>	<b>59,62</b>	51,39							
Y	<b>66,14</b>	<b>65,43</b>	<b>63,78</b>	<b>60,09</b>	51,45	50,44						
O	<b>73,12</b>	<b>71,58</b>	<b>68,77</b>	<b>66,78</b>	<b>56,88</b>	53,98	53,77					
C	<b>75,20</b>	<b>73,52</b>	<b>70,62</b>	<b>68,77</b>	<b>59,45</b>	<b>55,76</b>	<b>55,71</b>	52,32				
B	<b>78,15</b>	<b>76,37</b>	<b>73,42</b>	<b>72,06</b>	<b>63,23</b>	<b>59,23</b>	<b>59,14</b>	<b>56,27</b>	53,98			
R	<b>87,85</b>	<b>85,86</b>	<b>82,82</b>	<b>82,55</b>	<b>76,89</b>	<b>69,88</b>	<b>69,98</b>	<b>69,51</b>	<b>67,51</b>	<b>63,08</b>		
G	<b>91,66</b>	<b>89,89</b>	<b>87,21</b>	<b>87,10</b>	<b>83,71</b>	<b>76,22</b>	<b>76,15</b>	<b>77,12</b>	<b>75,43</b>	<b>71,17</b>	<b>59,80</b>	
T	<b>89,33</b>	<b>88,65</b>	<b>87,38</b>	<b>86,61</b>	<b>83,75</b>	<b>80,37</b>	<b>79,67</b>	<b>80,77</b>	<b>79,63</b>	<b>76,28</b>	<b>70,11</b>	<b>62,71</b>

Valores resaltados y en itálica indican diferencias significativas para  $p \leq 0,05$ .

Al utilizar el índice R para ordenamiento,  $R_{MAT}$ , se obtiene una variación en el orden del agrado con respecto al ANDEVA, en el cual se da la inversión de la posición de los productos Y y P; sin embargo ninguno de los dos métodos de análisis detectó diferencias entre estos productos. Con este índice se encontraron variaciones con el ANDEVA en cuanto a las diferencias encontradas entre otros productos; es decir, el índice  $R_{MAT}$  detectó diferencia entre los productos A-V, W-V, F-O, F-C, Y-C y P-C, las cuales no fueron demostradas con el análisis tradicional de los resultados de la escala hedónica de 9 puntos.

Lo anterior se asoció a los inconvenientes mencionados sobre el análisis tradicional de los datos mediante estadística paramétrica y evidenció que la estadística no paramétrica resultó ser más precisa al presentar menos posibilidades de incurrir en errores por el trato que se le da a los datos.

#### **5.2.2.2. $R_{JB}$**

En el Cuadro XX se presentan los resultados del índice  $R_{JB}$  para ordenamiento, los cuales se interpretan de la misma manera que el índice  $R_{MAT}$ . Se muestra que mediante este método se obtuvo un ordenamiento diferente de los productos de acuerdo al agrado que se detectó con el análisis tradicional, y se evidenció el siguiente orden de agrado, de mayor a menor: Z, W, A y V. Se observa que las variaciones se dieron en los tres productos de mayor agrado; sin embargo, se obtuvo el mismo resultado para los productos A-W, A-Z, W-Z y viceversa, es decir, no se detectaron diferencias significativas para estos productos con el análisis tradicional ni con el índice  $R_{JB}$  al analizar los productos de mayor agrado.

**Cuadro XX.** Índice  $R_{JB}$  para ordenamiento obtenido para trece productos mediante la utilización de la escala hedónica de 9 puntos.

Producto	Z	W	A	V	F	Y	P	O	C	B	R	G
W	50,50											
A	53,25	53,25										
V	55,25	53,50	53,25									
F	<b>63,00</b>	<b>72,25</b>	<b>73,50</b>	<b>62,00</b>								
Y	<b>68,50</b>	<b>63,00</b>	<b>62,25</b>	<b>60,50</b>	51,75							
P	<b>63,75</b>	<b>63,00</b>	<b>69,75</b>	<b>60,75</b>	50,50	50,50						
O	<b>69,75</b>	<b>71,25</b>	<b>76,00</b>	<b>64,25</b>	<b>59,25</b>	54,50	53,25					
C	<b>67,75</b>	<b>75,75</b>	<b>77,25</b>	<b>68,00</b>	<b>62,00</b>	<b>56,00</b>	<b>56,50</b>	51,25				
B	<b>69,75</b>	<b>75,75</b>	<b>78,75</b>	<b>71,25</b>	<b>65,00</b>	<b>58,00</b>	<b>61,00</b>	<b>56,50</b>	<b>57,25</b>			
R	<b>81,75</b>	<b>84,75</b>	<b>86,25</b>	<b>85,00</b>	<b>77,75</b>	<b>72,25</b>	<b>67,50</b>	<b>69,25</b>	<b>70,25</b>	<b>65,00</b>		
G	<b>81,00</b>	<b>91,00</b>	<b>93,00</b>	<b>84,00</b>	<b>88,25</b>	<b>70,00</b>	<b>81,00</b>	<b>88,50</b>	<b>73,50</b>	<b>68,00</b>	<b>57,75</b>	
T	<b>90,25</b>	<b>87,50</b>	<b>89,25</b>	<b>86,75</b>	<b>82,00</b>	<b>81,50</b>	<b>80,50</b>	<b>75,00</b>	<b>77,25</b>	<b>79,50</b>	<b>72,75</b>	<b>64,50</b>

Valores resaltados y en *itálica* indican diferencias significativas con  $p \leq 0,05$ .

Al comparar los resultados del análisis tradicional con los obtenidos con el índice  $R_{JB}$  se observa que este índice detectó diferencias que el ANDEVA no encontró. Estas variaciones se presentaron entre los productos F-O, F-C, Y-C, P-C y C-B.

Lo anterior permite ver que la estadística no paramétrica fue superior a la estadística paramétrica en el análisis de los resultados obtenidos con la escala hedónica de 9 puntos cuando se hace evaluación de más muestras que categorías. Esto a causa de que el índice R (estadística no paramétrica) no se vio afectado por criterios de decisión del uso de la escala y se basa en el comportamiento humano en lugar de estimaciones numéricas imperfectas obtenidas con una escala que funciona para mediciones ordinales en lugar de intervalo (Bi, 2006).

Así mismo, al utilizar el índice R se hizo uso de la estrategia cognitiva relativa de Mellers, es decir, se asume que el valor que se le da a un producto depende de los puntajes dados a los demás productos, y se tiene claro los límites y distancias entre los rangos (Feng & O'Mahony, 2017).

Se evidencia que las dos maneras de obtención del índice R ( $R_{MAT}$  y  $R_{JB}$ ), las cuales se compararán en el presente documento, permiten detectar mayores diferencias significativas que el ANDEVA/Fischer LSD; es decir, fue más potente (Cuadro XXX).

### 5.3. Análisis estadístico paramétrico y no paramétrico de menos muestras que categorías

#### 5.3.1. Análisis tradicional

El Cuadro XXI muestra los resultados obtenidos con el análisis de varianza, valor F y probabilidad asociada, del agrado de 5 muestras evaluadas mediante la escala hedónica de 9 puntos.

**Cuadro XXI.** Análisis de varianza para determinar diferencias en el agrado de 5 muestras de gomitas mediante la escala hedónica de 9 puntos.

F	Pr > F
54,626	< 0,0001

Se aprecia que el agrado generó un valor de probabilidad menor a 0,0001, es decir, había diferencia significativa en el agrado de al menos dos de las muestras mediante la escala hedónica de 9 puntos.

En el Cuadro XXII se evidencia el orden de preferencia obtenido para los productos de forma decreciente de izquierda a derecha (P-W-S-D-T), así como las diferencias entre los productos.

**Cuadro XXII.** Resultados de la comparación de medias de las calificaciones dadas por los consumidores a las cinco muestras analizadas.

Producto	P	W	S	D	T
Promedio	7,07 <sup>a</sup>	6,88 <sup>ab</sup>	6,58 <sup>bc</sup>	6,45 <sup>c</sup>	4,59 <sup>d</sup>

Letras diferentes indican diferencias significativas con  $p \leq 0,05$

En los primeros cuatro productos existió la tendencia que los dos productos consecutivos en agrado poseían el mismo nivel de satisfacción, es decir, no fueron diferentes estadísticamente. Así mismo, se apreció que el producto T era el que tenía menor agrado y era significativamente diferente a los demás productos.

Al utilizar un número de muestras menor al tamaño de la escala se permitió una mejor diferenciación de los productos debido a que había espacio suficiente en la escala hedónica verbal de 9 puntos para que se discriminaran únicamente 5 estímulos sin que existieran demasiados vínculos (Lim, 2011; Pimentel *et al.*, 2016; O'Mahony, 2017), así mismo el panelista logra ubicar el producto en el sitio que considere, clasificando una o más muestras en una misma categoría (Wichchukit & O'Mahony, 2015).

### 5.3.2. Índice R para ordenamiento

#### 5.3.2.1. $R_{MAT}$

Los valores del índice  $R_{MAT}$  para cinco muestras se presentan a continuación:

**Cuadro XXIII.** Índice R para ordenamiento ( $R_{MAT}$ ) obtenido para cinco productos mediante la utilización de la escala hedónica de 9 puntos

<b>Primero</b> <b>Segundo</b>	<b>P</b>	<b>W</b>	<b>S</b>	<b>D</b>
<b>W</b>	52,33			
<b>S</b>	<b>57,72</b>	54,84		
<b>D</b>	<b>61,90</b>	<b>58,83</b>	54,05	
<b>T</b>	<b>84,69</b>	<b>81,25</b>	<b>80,97</b>	<b>78,76</b>

Valores resaltados y en itálica indican diferencias significativas.

Para interpretar el Cuadro anterior se debe tomar en cuenta que, al igual que con el análisis de 13 muestras, un valor de 50% en el índice R significa que no hay diferencias mientras que entre más cercano a 100% sea este índice, mayor diferencia habría en cuanto a agrado, de manera que si el valor del índice R aumenta se le asigna un mayor agrado al producto que está en la primera fila; con respecto al producto de la fila inferior.

Se observa que no hubo diferencias significativas entre los productos P y W, entre W y S; y entre S y D. Entre los demás productos hubo diferencias significativas. El orden de agrado de los productos (P-W-S-D-T) obtenido con el

índice R para ordenamiento ( $R_{MAT}$ ) que se presenta en el Cuadro XXIII coincidió con el obtenido mediante el análisis de datos tradicional expuesto en el Cuadro XXII.

En cuanto a las diferencias detectadas con índice  $R_{MAT}$  se aprecia que concuerdan con lo obtenido con el método tradicional, con lo cual se confirmó que no existió diferencia entre la utilización de la estadística paramétrica y la no paramétrica en este caso. Lo anterior se sustenta en la afirmación de que el índice R es uno de los métodos más potentes en la estadística no paramétrica (Bi, 2006) y que al utilizar menos muestras que categorías se presentaron menos vínculos al utilizar la estadística paramétrica (ANDEVA/Fischer LSD) en comparación con el análisis de 13 muestras y, por consiguiente, no se presentaron los inconvenientes que se resaltaron en el análisis de más muestras que categorías.

### 5.3.2.2. $R_{JB}$

El análisis utilizando el índice R de John Brown, estadística no paramétrica, se presenta seguidamente.

**Cuadro XXIV.** Índice  $R_{JB}$  para ordenamiento obtenido para cinco productos mediante la utilización de la escala hedónica de 9 puntos

Primero \ Segundo	P	W	S	D
W	51,50			
S	<b>59,25</b>	53,50		
D	<b>60,50</b>	<b>59,75</b>	<b>56,00</b>	
T	<b>84,75</b>	<b>79,75</b>	<b>79,25</b>	<b>80,50</b>

Valores resaltados y en itálica indican diferencias significativas.

Se observa que no hubo diferencias significativas entre los productos P y W ni entre W y S. Al obtener el índice R para ordenamiento,  $R_{JB}$ , se obtuvo el mismo orden de agrado que con el índice  $R_{MAT}$  y el análisis tradicional (P-W-S-D-T). Se aprecia que el producto S no fue diferente ( $p > 0,05$ ) en agrado al producto D, mediante el análisis tradicional (Cuadro XXII) y el índice  $R_{MAT}$  (Cuadro XVIII);

no obstante, mediante el  $R_{JB}$  se determinó una diferencia significativa ( $p \leq 0,05$ ), entre los productos S y D, que no fue detectada por los últimos dos métodos mencionados.

Al analizar menos muestras que categorías se determinó que el índice  $R_{JB}$  para ordenamiento (estadística no paramétrica) es más potente que la estadística paramétrica (ANDEVA/Fischer LSD) (Cuadro XXXI). Lo anterior no se esperaba debido a que al ser menos estímulos que categorías no se incurre en errores que si se presentan al utilizar mayor cantidad de muestras; lo cual evidencia que el emplear la estadística no paramétrica puede evitar problemas de análisis y conclusiones finales (O`Mahony, 2017).

#### 5.4. Comparación del índice $R_{JB}$ y $R_{MAT}$ para dos tamaños de muestras

Los resultados obtenidos para los cinco productos mediante el método tradicional de análisis coincidieron con los obtenidos con el índice  $R_{MAT}$  y el  $R_{JB}$  en el orden del agrado, sin embargo, presentaron variaciones en la detección de diferencias, siendo el índice  $R_{JB}$  más sensible para encontrar diferencias.

Tomando las diferencias significativas encontradas mediante al menos uno de los índices,  $R_{JB}$  y  $R_{MAT}$ , para 13 muestras (Cuadro XIX y Cuadro XX) y 5 muestras (Cuadro XXIII y Cuadro XXIV) se compararon los índices obtenidos con ambos métodos y se contabilizó el número de veces que  $R_{JB}$  fue mayor a  $R_{MAT}$  y viceversa. Lo anterior se presenta en el siguiente Cuadro:

**Cuadro XXV.** Número de veces que  $R_{JB} > R_{MAT}$  y  $R_{MAT} > R_{JB}$  a partir de las diferencias significativas detectadas en al menos uno de los índices R para ordenamiento, para dos tamaños de muestras distintas

Número de muestras	Total de pares con diferencias significativas	Valor de diferencias significativas donde $R_{JB} > R_{MAT}$	Valor de diferencias significativas donde $R_{MAT} > R_{JB}$
5	8	5	3
13	68	37	31

El Cuadro XXV muestra que para ambos tamaños de muestras se obtuvo mayor cantidad de valores de  $R_{JB}$  superiores a  $R_{MAT}$ . Lo anterior sugiere que la capacidad para detectar diferencias fue superior para el primero de éstos, ya que al generar valores mayores hace mayor diferenciación entre dos productos.

Las variaciones entre ambos índices ha sido estudiada tanto para pruebas de diferencia (O'Mahony *et al.*, 1980) como para pruebas de agrado con escala hedónica, en las cuales se ha evidenciado que en algunas ocasiones el  $R_{JB}$  es mayor y en otras lo es el  $R_{MAT}$  (O'Mahony, 2017).

Brown (1974) predijo estas diferencias, y las asoció con el método del cálculo del índice, ya que al elaborar la matriz con las etiquetas para la obtención del  $R_{MAT}$  se adopta una metodología de cálculo en la cual se comprimen los datos (O'Mahony *et al.*, 1980) y se crea una mayor vinculación de éstos, es decir, al unirlos en una matriz se crean lazos aparentes entre las muestras que generan un valor mayor del índice  $R_{MAT}$ . (O'Mahony, 2017). Los valores del  $R_{JB}$  son generalmente mayores debido a que al no realizar la construcción de una matriz y al ser obtenido de manera manual se evita la compresión de los datos y los vínculos aparentes (O'Mahony, 2017).

Aunque la evaluación con 5 muestras fue de sabor y la de 13 productos fue visual, al comparar la ubicación en la escala hedónica de los 3 productos que se repitieron en los dos estudios (Goma Pino (P), Watermelon (W) y Teeth (T)), se encontró que la goma W estuvo en segundo lugar, la P en posición intermedia y la T en último lugar cuando se evaluaron 5 o 13 productos y con análisis estadístico paramétrico o no paramétrico, lo cual permite apreciar que se mantuvo el orden de agrado de los productos.

La variación entre la prueba de 5 muestras (sabor) y la prueba de 13 muestras (visual) se debe a la fatiga que se ocasiona en el panelista al tener que probar 13 muestras. Debido a lo anterior se recomienda al evaluar sabor utilizar una cantidad de productos menor al número de categorías, para evitar el cansancio del panelista y favorecer una buena evaluación de las muestras (Hernandez, 2005).

## 6. CONCLUSIONES

A partir de los resultados obtenidos en la investigación se concluye lo siguiente:

- ✓ La escala elaborada por los consumidores no coincide con la traducción directa de la escala hedónica de 9 puntos en inglés.
- ✓ La estadística no paramétrica resultó más potente que la estadística paramétrica para la detección de diferencias entre productos cuando se usa un número de muestras mayor al número de categorías presentes en la escala hedónica de nueve puntos.
- ✓ Al utilizar estadística no paramétrica se obtuvo mayor número de muestras con diferencias significativas, en comparación con la estadística paramétrica, al utilizar un tamaño de muestras mayor y uno menor que el número de categorías de la escala hedónica de 9 puntos.
- ✓ El índice  $R_{JB}$  permitió detectar diferencias mayores entre las muestras, en la mayoría de las ocasiones que se presentaron diferencias significativas entre productos, en comparación con el índice  $R_{MAT}$ , para dos tamaños de muestras distintas (5 y 13); es decir, el  $R_{JB}$  fue más potente.

## **7. RECOMENDACIONES**

A partir de los resultados obtenidos en esta investigación se recomienda:

- Utilizar la estadística no paramétrica, Índice R para ordenamiento, para el análisis de los datos de agrado obtenidos con la escala hedónica de 9 puntos, independientemente del número de muestras a evaluar.
- Cuando se requiera utilizar la escala hedónica de 9 puntos en español, se recomienda verificar que los consumidores organizan las categorías en la forma que se estableció en este estudio antes de que la use el participante, de manera que esté de acuerdo con la escala. Si no estuviera de acuerdo con la escala, ese participante debe abstenerse de hacer la evaluación.

## 8. BIBLIOGRAFÍA

- Angulo, O. & O'Mahony, M. 2009. Las pruebas de preferencia en alimentos son más complejas de lo imaginado. *Interciencia*, **34**, 177–181.
- Basker, D. 1998. Critical Values of differences among rank sums for multiple comparison. *Journal of Food Technology*, **42**, 79–84.
- Bi, J. 2006. Statistical analyses for R-index. *Journal of Sensory Studies*, **21**, 584–600.
- Bi, J. & O'Mahony, M. 2007. Updated and extended table for testing the significance of the r-index. *Journal of Sensory Studies*, **22**, 713–720.
- Brody, A.L. & Lord, J.B. 2007. *Developing new food products for a changing marketplace*. CRC Press.
- Brown, J. 1974. Recognition assessed by rating and ranking. *British Journal of Psychology*, **65**, 13–22.
- Carpenter, R.P., Lyon, D.H. & Hasdell, T.A. 2012. *Guidelines for Sensory Analysis in Food Product Development and Quality Control*. Springer US.
- Castañeda, C. 2013. *Comparación de la escala hedónica de nueve puntos con la escala hedónica general de magnitud (gLMS) utilizada por personas de dos regiones de América Latina*.
- Clark, S., Costello, M., Drake, M. & Bodyfelt, F. 2009. *The Sensory Evaluation of Dairy Products*. New York, NY: Springer US.
- Cliff, M.A., King, M.C., Scaman, C. & Edwards, B.J. 1997. Evaluation of R-indices for preference testing of apple juices. *Food Quality and Preference*, **8**, 241–246.
- Curia, A.V., Hough, G., Martínez, M.C. & Margalef, M.I. 2001. How Argentine consumers understand the Spanish translation of the 9-point hedonic scale. *Food Quality and Preference*, **12**, 217–221.
- Drake, M. 2007. Invited Review: Sensory Analysis of Dairy Foods. *Journal of Dairy Science*, **90**, 4925–4937.
- Espinosa, P., Villacrés, E., Bautista, C. & Espin, S. 1998. *El uso del análisis sensorial para medir la aceptación de clones promisorios de papa: una experiencia ecuatoriana*. Ediciones Abya-Yala.
- Feng, Y. & O'Mahony, M. 2017. Comparison between American and Chinese consumers in the use of verbal and numerical 9-point hedonic scales and R-Index ranking for food and personal products. *Food Quality and Preference*,

60, 138–144.

- Feng, Y.H., Gutiérrez-Salomón, A.L., Angulo, O., O'Mahony, M. & Wichchukit, S. 2015. Data from “words only” and “numbers only” 9-point hedonic scales are not interchangeable for serial monadic as well as rank-rating protocols: Aspects of memory and culture. *Food Quality and Preference*, **41**, 12–19.
- González, V., Carlos, R., Mauriz, R., Sanmartín, C., Sergio, F., Plana, V., Covadonga, T. & Rey, R.-M. 2014. Introducción al análisis sensorial Estudio hedónico del pan en el IES Mugaros Categoría: Bachillerato y ciclos formativos.
- Hernandez, E. 2005. *Evaluacion sensorial*. Bogotá: UNAD.
- Hough, G., Wakeling, I., Mucci, A., Chambers IV, E., Gallardo, I.M. & Alves, L.R. 2006. Number of consumers necessary for sensory acceptability tests. *Food Quality and Preference*, **17**, 522–526.
- Ibáñez Moya, F.C. & Barcina Angulo, Y. 2001. *Análisis sensorial de alimentos: métodos y aplicaciones*. Springer.
- Jaeger, S. & Cardello, A. 2009. Direct and indirect hedonic scaling methods: A comparison of the labeled affective magnitude (LAM) scale and best–worst scaling. *Food Quality and Preference*, **20**, 249–258.
- Jones, L. V., Peryam, D.R. & Thurstone, L.L. 1955. Development of a scale for measuring soldiers' food preferences. *Journal of Food Science*, **20**, 512–520.
- Lawless, H.T. & Heymann, H. 2013. *Sensory evaluation of food: principles and practices*. Springer.
- Lawless, H.T., Popper, R. & Kroll, B.J. 2010. A comparison of the labeled magnitude (LAM) scale, an 11-point category scale and the traditional 9-point hedonic scale. *Food Quality and Preference*, **21**, 4–12.
- Lee, H.S. & Hout, D. Van. 2009. Quantification of sensory and food quality: The R-index analysis. *Journal of Food Science*, **74**, 57–64.
- Lee, H.S. & O'Mahony, M. 2005. Sensory evaluation and marketing: Measurement of a consumer concept. *Food Quality and Preference*, **16**, 227–235.
- Lim, J. 2011. Hedonic scaling: A review of methods and theory. *Food Quality and Preference*, **22**, 733–747.
- Lim, J. & Fujimaru, T. 2010. Evaluation of the Labeled Hedonic Scale under different experimental conditions. *Food Quality and Preference*, **21**, 521–530.

- Meilgaard, M., Civille, G.V. & Carr, B.T. 2007. *Sensory evaluation techniques*. Taylor & Francis.
- Mellers, B.A. 1983a. Evidence against “absolute” scaling. *Perception & Psychophysics*, **33**, 523–526.
- Mellers, B.A. 1983b. Reply to Zwislocki’s views on “absolute” scaling. *Perception & Psychophysics*, **34**, 405–408.
- Mondino, M.C. & Ferratto, J. 2006. El análisis sensorial, una herramienta para la evaluación de la calidad desde el consumidor - Facultad de Ciencias Agrarias - UNR [Internet] URL <http://www.fcagr.unr.edu.ar/Extension/Agromensajes/18/7AM18.htm>. Accessed 15/11/2017.
- Nicolas, L., Marquilly, C. & O`Mahony, M. 2010. The 9-point hedonic scale: Are words and numbers compatible? *Food Quality and Preference*, **21**, 1008–1015.
- O`Mahony, M., Garske, S. & Klapman, K. 1980. Rating and Ranking Procedures for Short Cut Signal Detection Multiple Difference Tests. *Journal of Food Science*, **45**, 392–393.
- O`Mahony, M., Park, H., Park, J.Y. & Kim, K.O.K. 2004. Comparison of the statistical analysis of hedonic data using analysis of variance and multiple comparisons versus an R-index analysis of the ranked data. *Journal of Sensory Studies*, **19**, 519–529.
- O`Mahony, M. 2017. *Comunicación Personal*.
- Park, H., O`Mahony, M. & Kim, K.O. 2007. A comparison of the discriminating power of anova and r-index analyses of hedonic data for various products and experimental protocols. *Journal of Sensory Studies*, **22**, 281–292.
- Peryam, D. & Pilgrim, F. 1957. Hedonic scale method of measuring food preferences. *Food Technology*, **11**.
- Peryam, D., Polemis, B., Kamen, J., Eindhoven, J. & Pilgrim, F. 1960. Food Preferences of men in the U.S. armed forces. *Department of the Army Quartermaster Research and Engineering Command-Quartermaster Food and Container Institute for the Armed Forces*.
- Peryam, D.R. & Girardot, N.F. 1952. Advanced taste test method”. *Food Engineering*,. *Food Engineering*, **24**, 58–61.
- Pilgrim, F. & Peryam, D. 1996. *Sensory Testing Methods: A Manual*. Second.

- Philadelphia: ASTM International.
- Pimentel, T.C., Gomes da Cruz, A. & Deliza, R. 2016. Sensory Evaluation: Sensory Rating and Scoring Methods. In: *Encyclopedia of Food and Health*. Pp. 744–749. Elsevier.
- Ramírez, J.S. 2012. Análisis sensorial: pruebas orientadas al consumidor. *ReCiTeIA*, **12**, 83–102.
- Robinson, K.M., Klein, B.P. & Lee, S.Y. 2004. Utilizing the R-Index Measure for Threshold Testing in Model Soy Isoflavone Solutions. *Journal of Food Science*, **69**, SNQ1-SNQ4.
- Rosas Nexticapa, M., Angulo, O. & O'Mahony, M. 2005. How well does the 9-point hedonic scale predict purchase frequency? *Journal of Sensory Studies*, **20**, 313–331.
- Sancho, J., Bota, E., Castro, J.J. de & Puig, E. 1999. *Introducción al análisis sensorial de los alimentos*. Edicions Universitat de Barcelona.
- Schutz, H.G. & Cardello, A. V. 2001. A labeled affective magnitude (lam) scale for assessing food liking/disliking. *Journal of Sensory Studies*, **16**, 117–159.
- Simone, M. & Pangborn, R. 1957. Consumer acceptance methodology; one vs two samples. *Food Technology*, **11**, 25–29.
- Varela, P., Beltrán, J. & Fiszman, S. 2014. An alternative way to uncover drivers of coffee liking: Preference mapping based on consumers' preference ranking and open comments. *Food Quality and Preference*, **32**, 152–159.
- Villanueva, N.D.M., Petenate, A.J. & Silva, M.A.A.P. Da. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. *Food Quality and Preference*, **16**, 691–703.
- Watts, B.M., Yllimaki, G., Jeffreys, L. & Elias, L. 1989. *Basic sensory methods for food evaluation*. Canada: The Centre.
- Wichchukit, S. & O'Mahony, M. 2015. The 9-point hedonic scale and hedonic ranking in food science: Some reappraisals and alternatives. *Journal of the Science of Food and Agriculture*, **95**, 2167–2178.
- Xia, Y., Zhong, F. & O'Mahony, M. 2015. Is the Discrepancy between Numbers Derived from Verbal and Numerical Protocols for 9-Point Hedonic Scales an Artifact of Product Choice? *Journal of Sensory Studies*, **30**, 269–279.
- Yao, E., Lim, J., Tamaki, K., Ishii, R., Kim, K.-O. & O'Mahony, M. 2003. Structured and unstructured 9-point hedonic scales: a cross cultural study with

- american, japanese and korean consumers. *Journal of Sensory Studies*, **18**, 115–139.
- Yeh, L., Kim, K., Chompreeda, P., Rimkeeree, H., Yau, N.J.. & Lundahl, D.. 1998. Comparison in Use of the 9-Point Hedonic Scale between Americans, Chinese, Koreans, and Thai. *Food Quality and Preference*, **9**, 413–419.
- Zwislocki, J.J. 1983. Absolute and other scales: Question of validity. *Perception & Psychophysics*, **33**, 593–594.
- Zwislocki, J.J. & Goodman, D.A. 1980. Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, **28**, 28–38.

## 9. ANEXOS

### 9.1. Hoja de recolección de datos para análisis de cinco muestras

Nombre: \_\_\_\_\_ Sexo F: \_\_\_ M: \_\_\_ Edad: \_\_\_

Me disgusta muchísimo	Me disgusta mucho	Me disgusta moderadamente	Me disgusta un poco	No me gusta ni me disgusta	Me gusta un poco	Me gusta moderadamente	Me gusta mucho	Me gusta muchísimo
DES:	DES:	DES:	DES:	DES:	DES:	DES:	DES:	DES:

DES: Desempate de productos

## 9.2. Matriz para obtención del índice $R_{MAT}$ para cinco muestras

**Cuadro XXVI.** Matriz para la obtención del índice  $R_{MAT}$  para un tamaño de muestras menor al número de categorías de la escala hedónica de 9 puntos.

Código	Posición								
	1	1,5	2	2,5	3	3,5	4	4,5	5
<b>P</b>	65	4	38	0	46	1	29	3	14
<b>W</b>	65	4	36	1	31	0	37	1	25
<b>S</b>	38	4	45	1	47	0	46	1	18
<b>D</b>	21	1	54	2	54	1	43	1	23
<b>T</b>	4	1	17	0	18	2	37	6	115

Los números en cada casilla corresponden al número de veces (consumidores) que se colocó el producto en cada posición en la escala.

### 9.3. Matriz para obtención del índice $R_{MAT}$ para trece muestras

**Cuadro XXVII.** Matriz para la obtención del índice  $R_{MAT}$  para un tamaño de muestras mayor al número de categorías de la escala hedónica de 9 puntos.

Producto	Posición																								
	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10	10,5	11	11,5	12	12,5	13
<b>A</b>	0	18	0	31	1	37	1	38	0	200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>R</b>	0	0	2	0	4	0	7	0	9	0	12	0	21	0	16	1	22	0	23	1	30	0	37	1	14
<b>T</b>	4	0	2	2	3	0	4	0	6	0	5	1	7	0	7	0	6	1	11	0	14	0	22	1	101
<b>G</b>	16	0	2	0	2	0	3	0	5	1	6	1	11	0	17	0	26	0	18	0	31	1	37	1	38
<b>C</b>	36	1	10	3	12	0	15	1	14	0	15	1	20	0	22	0	23	2	20	0	19	0	15	1	2
<b>B</b>	13	0	8	1	11	1	11	2	9	2	19	2	18	0	22	0	23	2	18	1	15	0	19	0	12
<b>Y</b>	39	0	18	1	28	0	11	0	11	0	15	2	11	1	12	1	12	2	18	0	16	0	12	0	13
<b>W</b>	9	2	33	0	12	0	25	0	22	0	21	2	10	0	10	0	7	0	9	0	5	0	5	0	1
<b>V</b>	23	1	25	2	33	2	28	1	20	1	13	1	11	1	11	0	9	1	11	2	3	0	9	0	2
<b>Z</b>	7	3	31	1	17	1	16	0	16	0	16	1	7	0	12	0	8	0	13	0	14	0	3	0	2
<b>F</b>	0	0	9	2	14	0	21	1	25	1	18	1	21	0	28	2	20	0	15	0	10	0	1	0	2
<b>P</b>	0	2	18	0	12	0	18	0	17	0	10	1	12	0	12	0	14	0	18	0	15	1	19	0	8
<b>O</b>	0	1	13	1	11	0	14	0	14	3	20	0	27	0	16	0	13	2	24	0	16	0	18	0	0

Los números en cada casilla corresponden al número de personas que colocaron el producto en esa posición en la escala

## 9.4. Muestra de cálculos del índice R para ordenamiento

### 9.4.1. $R_{JB}$

Se detalla la muestra del cálculo del índice  $R_{JB}$  para ordenamiento, utilizando los dos primeros productos en agrado de la evaluación de 5 muestras

**Cuadro XXVIII.** Posición de cada muestra de la evaluación de 5 productos mediante la escala hedónica de 9 puntos usando 20 jueces.

Consumidor	1	1,5	2	2,5	3	3,5	4	4,5	5
1	P		D		S		T		W
2	S		D		W		P		T
3	P		D		S		W		T
4	W		S		D		T		P
5	D		S		W		T		P
6	W		D		T		P		S
7	W		P		D		S		T
8	P		D		T		W		S
9	W		S		P		T		D
10	P		S		T		W		D
11	W		P		S		T		D
12	T		P		D		S		W
13	S			W D			P		T
14	S		W		D		P		T
15	W		P		D		S		T
16	W		S		P		D		T
17	D		S		P		W		T
18	P		W		D		S		T
19	W		P		D		S		T
20	P		T		W		S		D

Las letras en cada casilla corresponden al código de la muestra que se colocó en esa categoría de la escala y el sombreado se fijó a la muestra W

solamente cuando estuvo antes que D. Se presentó doce veces antes W que D y la mitad del juez 13.

El siguiente ejemplo muestra la preferencia de W sobre D, de acuerdo a 20 consumidores:

$$W \rightarrow D = \frac{X}{N} * 100$$

Donde:

P<sub>1</sub>: Producto 1 (mayor agrado)

P<sub>2</sub>: Producto 2 (menor agrado)

X: Cantidad de veces que un P<sub>1</sub> presenta mayor agrado con respecto a P<sub>2</sub>

N: Número total de jueces

$$W \rightarrow D = \frac{12,5}{20} * 100 = 62,50\%$$

### 9.4.2. $R_{MAT}$

Se detalla la muestra del cálculo del índice  $R_{MAT}$  para ordenamiento, utilizando los dos primeros productos en agrado de la evaluación de 5 muestras

**Cuadro XXIX.** Número de veces que cada producto se encuentra en determinada posición, obtenido mediante la evaluación de menos muestras que categorías mediante la escala hedónica de 9 puntos.

Producto	1	1,5	2	2,5	3	3,5	4	4,5	5	m
GOMA PINO (P)	65	4	38	0	46	1	29	3	14	200
WATERMELON SLICES (W)	65	4	36	1	31	0	37	1	25	200

Los valores 1,5; 2,5; 3,5 y 4,5 se obtienen cuando dos productos se ubican en la misma categoría y el panelista establece un empate en el agrado de los productos, en este caso se le asigna el valor intermedio a cada producto.

El siguiente ejemplo muestra la preferencia de P sobre W, de acuerdo a 200 consumidores:

$$\begin{aligned}
 P_W = & P1 * (W2 + W3 + W4 + W5 + W6 + W7 + W8 + W9) + P2 \\
 & * (W3 + W4 + W5 + W6 + W7 + W8 + W9) + P3 \\
 & * (W4 + W5 + W6 + W7 + W8 + W9) + P4 \\
 & * (W5 + W6 + W7 + W8 + W9) + P5 * (W6 + W7 + W8 + W9) + P6 \\
 & * (W7 + W8 + W9) + P7 * (W8 + W9) + P8 * (W9) \\
 & + \frac{1}{2} (P1 * W1 + P2 * W2 + P3 * W3 + P4 * W4 + P5 * W5 + P6 * W6 \\
 & + P7 * W7 + P8 * W8 + P9 * W9)
 \end{aligned}$$

$$\begin{aligned}
 P_W = & 65 * (4 + 36 + 1 + 31 + 0 + 37 + 1 + 25) + 4 \\
 & * (36 + 1 + 31 + 0 + 37 + 1 + 25) + 38 * (1 + 31 + 0 + 37 + 1 + 25) \\
 & + 0 * (31 + 0 + 37 + 1 + 25) + 46 * (0 + 37 + 1 + 25) + 1 \\
 & * (37 + 1 + 25) + 29 * (1 + 25) + 3 * 25 \\
 & + \frac{1}{2} (65 * 65 + 4 * 4 + 38 * 36 + 0 * 1 + 46 * 31 + 1 * 0 + 29 * 37 + 3 \\
 & * 1 + 14 * 25)
 \end{aligned}$$

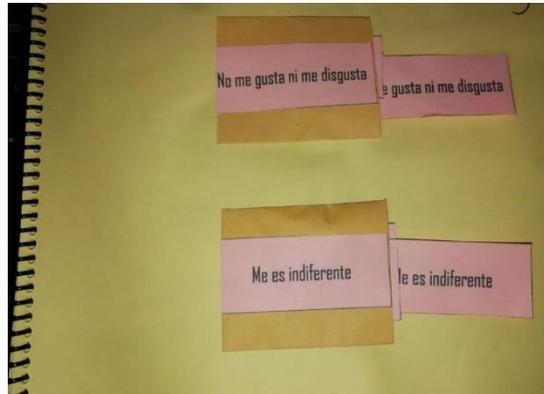
$$P_W = 20\,929,5$$

$$\%P_W = \frac{P_W}{N * N} * 100$$

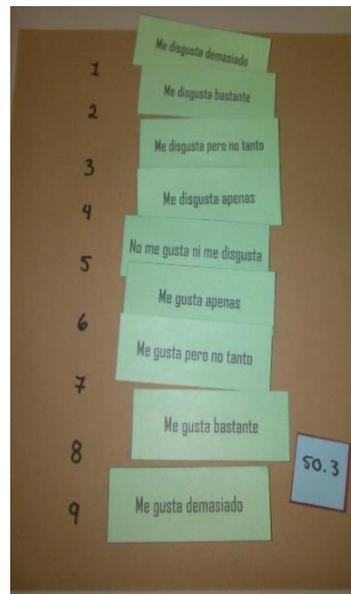
$$\%P_W = \frac{20\,929,5}{200 * 200} * 100 = 52,32\%$$

### 9.5. Herramientas para la construcción de la escala hedónica de 9 puntos traducida al español.

Se presentan las herramientas utilizadas en la construcción de la escala hedónica de 9 puntos traducida al español.



**Figura 2.** Tarjetas utilizadas en la construcción de la escala hedónica de 9 puntos traducida al español.



**Figura 3.** Ejemplo de las categorías elaboradas por una muestra de la población costarricense.

### 9.6. Comparación de resultados de análisis paramétrico y análisis no paramétrico.

Los siguientes cuadros muestran las comparaciones de los resultados con los tres métodos de obtención: Andeva,  $R_{MAT}$  y  $R_{JB}$ , para dos tamaños de muestras.

**Cuadro XXX.** Comparación de las diferencias encontradas para las trece muestras analizadas con estadística paramétrica y los índices R.

Producto	A	W	Z	V	F	Y	P	O	C	B	R	G	T
<b>Promedio (Paramétrico)</b>	7,21 <sup>a</sup>	7,16 <sup>a</sup>	7,06 <sup>a</sup>	6,91 <sup>a</sup>	6,19 <sup>b</sup>	6,16 <sup>b</sup>	6,09 <sup>b</sup>	5,93 <sup>b</sup>	5,85 <sup>bc</sup>	5,54 <sup>c</sup>	4,88 <sup>d</sup>	4,45 <sup>e</sup>	3,84 <sup>f</sup>
<b>Producto <math>R_{JB}</math></b>	Z	W	A	V	F	Y	P	O	C	B	R	G	T
<b>No paramétrico, <math>R_{JB}</math></b>	a	a	a	A	b	bc	bc	c	D	E	f	g	h
<b>Producto <math>R_{MAT}</math></b>	A	W	Z	V	F	P	Y	O	C	B	R	G	T
<b>No paramétrico, <math>R_{MAT}</math></b>	a	a	ab	B	c	cd	cd	de	Ef	F	g	h	i

Letras diferentes indican diferencias significativas con  $p \leq 0,05$

**Cuadro XXXI.** Comparación de las diferencias encontradas para las cinco muestras analizadas con estadística paramétrica y los índices R.

Producto	P	W	S	D	T
<b>Promedio (Paramétrico)</b>	7,07 <sup>a</sup>	6,88 <sup>ab</sup>	6,58 <sup>bc</sup>	6,45 <sup>c</sup>	4,59 <sup>d</sup>
<b>Producto <math>R_{JB}</math></b>	P	W	S	D	T
<b>No paramétrico, <math>R_{JB}</math></b>	a	ab	B	c	d
<b>Producto <math>R_{MAT}</math></b>	P	W	S	D	T
<b>No paramétrico, <math>R_{MAT}</math></b>	a	ab	Bc	c	d

Letras diferentes indican diferencias significativas con  $p \leq 0,05$