



UNIVERSIDADE DA CORUÑA

Universidade de Vigo

Trabajo Fin de Máster

**Análisis e implementación de técnicas de análisis sensorial de
discriminación en R**

Melissa Beras Caraballo

Máster en Técnicas Estadísticas

Curso 2022-2023

Propuesta de Trabajo Fin de Máster

Título en galego: Análisis e implementación de técnicas de análisis sensorial de discriminación en R.
Título en español: Análisis e implementación de técnicas de análisis sensorial de discriminación en R
English title: Analysis and implementation of discriminative sensory analysis techniques in R.
Modalidad: Modalidad B.
Autor: Melissa Beras Caraballo, Universidade da Coruña.
Director: César Andrés Sánchez Sellero, Universidade de Santiago de Compostela.
Tutora: Maruxa Quiroga García, TasteLab S. L.
Breve resumen del trabajo: Análisis de los métodos de análisis sensorial y las pruebas estadísticas correspondientes para su implementación en R y recomendación de acuerdo con el tipo de producto a evaluar y la cantidad de recursos disponibles.
Recomendaciones:
Otras observaciones: Este trabajo fin de máster se enfoca en las técnicas estadísticas para el análisis de los resultados de las pruebas discriminativas, por lo que solo se definieron los conceptos básicos de las pruebas de análisis sensorial y se recomienda la revisión de las referencias bibliográficas para mayor entendimiento sobre el procedimiento a seguir en términos de presentación de las muestras y creación de la sala de cata para la sesión de evaluación.

Don César Andrés Sánchez Sello, profesor titular de la Universidad de Santiago de Compostela y Doña Maruxa Quiroga García, CEO de Tastelab SL, informan que el Trabajo Fin de Máster titulado

Análisis e implementación de técnicas de análisis sensorial discriminativo en R

fue realizado bajo su dirección por doña Melissa Beras Caraballo para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En A Coruña, a 13 de julio de 2023.

El director:

Don César Andrés Sánchez Sello

La tutora:

Doña Maruxa Quiroga García

La autora:

Doña Melissa Beras Caraballo

Declaración responsable. Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, Disposición 2978 del BOE núm. 48 de 2022), el/la autor/a declara que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas,. . .)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración,. . . sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.

Agradecimientos

El tiempo sigue pasando y está en nuestras manos aprovecharlo. Esta ha sido la mayor enseñanza de mis padres, Julio Cesar y Nidia, a quienes quiero agradecer por su constante apoyo en todos mis proyectos y por inspirarme a cada día dar lo mejor de mí. Ellos me han enseñado que, con esfuerzo y determinación, todo es posible.

A todo el equipo de Tastelab, quienes durante unos meses me aliviaron la carga que supone cursar un máster en técnicas estadísticas sin tener gran experiencia en el campo. El verlos trabajar en equipo y con tanto entusiasmo, me hizo recordar lo importante que será poder aportar los conocimientos del máster a las necesidades de la industria. Quiero agradecer especialmente a mi tutora Maruxa Quiroga, por darme la oportunidad de conocer sobre el análisis sensorial, por su confianza, empatía, paciencia y flexibilidad durante todo el proceso, pero sobre todo por recibirme con una sonrisa desde el primer día.

A mi director académico, César Sánchez, por su flexibilidad y disposición durante todo el proceso, por sus palabras duras pero que aterrizan y ayudan a ver las debilidades como oportunidades de mejora.

A mi querido César Adel y a mis hermanos Julio A, Julio E y Claudia, por siempre escucharme y estar presente durante todo este trayecto.

Tabla de contenidos

1	Análisis sensorial	10
1.1	Introducción	10
1.2	Métodos de prueba de análisis sensorial	11
	Pruebas discriminativas	11
	Pruebas descriptivas	11
	Pruebas afectivas	11
1.3	Normas de análisis sensorial	12
1.3.1	UNE-EN ISO 5492:2010: Análisis sensorial. Vocabulario	12
1.3.2	UNE-ISO 6658:2019: Análisis sensorial. Metodología. Guía general	14
2	Pruebas discriminativas	16
2.1	Métodos de selección forzada	17
2.1.1	Prueba triangular	17
2.1.2	Prueba tetrad	18
2.1.3	Prueba de elección forzada de 1 entre 2 alternativas (2-AFC)	18
2.1.4	Prueba de elección forzada de 1 entre 3 alternativas (3-AFC)	18
2.1.5	Prueba duo-trío	19
2.1.6	Prueba diferente del control (DFC)	19
2.1.7	Prueba dos de cinco	20
2.1.8	Prueba de clasificación	21
2.2	Métodos con sesgo de respuesta	21
2.2.1	Prueba igual – diferente	21
2.2.2	Prueba “A” – “No A”	22
2.3	Comparación y selección de las pruebas discriminativas	23
3	Técnicas estadísticas para pruebas discriminativas de diferencia	26
3.1	Prueba binomial	27
3.2	Prueba T de Student	34
3.3	Análisis de la varianza (ANOVA)	36
3.4	Prueba de los rangos con signo de Wilcoxon	39
3.5	Prueba de Friedman	40
3.6	Prueba de chi cuadrado (χ^2)	41
3.6.1	Prueba de chi cuadrado χ^2 de Pearson para homogeneidad	43

3.6.2	Prueba de chi cuadrado χ^2 de Pearson para independencia	45
3.6.3	Prueba de chi cuadrado χ^2 de McNemar para proporción correlacionada	46
4	Análisis de la potencia estadística.....	48
4.1	Cálculo de tamaño muestral mediante fórmulas estadísticas.....	49
4.1.1	Prueba Binomial.....	49
4.1.2	Prueba T de Student.....	51
4.1.3	ANOVA	53
4.1.4	Prueba χ^2 de Pearson	55
4.2	Cálculo de tamaño muestral mediante técnicas de simulación	57
4.2.1	Prueba Binomial.....	57
4.2.2	Prueba T de Student.....	59
4.2.3	Prueba ANOVA.....	60
4.2.4	Prueba de Wilcoxon	61
4.2.5	Prueba de Friedman.....	62
4.3	Conclusiones	63
5	Diseño experimental.....	64
6	Implementación de las pruebas.....	66
6.1	Función “prueba_binomial()”.	66
6.2	Función “prueba_DFC()”	67
6.3	Función “prueba_chisqA()”	67
6.4	Función “prueba_chisqID()”	68
7	Simulación de datos y ejecución de las pruebas.....	70
7.1	Prueba binomial.....	70
7.2	Prueba DFC.....	71
7.3	Prueba “A” –“No A” / “Igual” -“Diferente”	72
	Conclusiones	74
	Bibliografía	75
	Anexos.....	77
	Anexo 1: Análisis de la potencia	77
	Análisis de la potencia mediante fórmulas estadísticas	77
	Análisis de la potencia mediante simulación estadística.....	78
	Prueba Binomial.....	78
	Prueba T de Student	79

Prueba ANOVA	80
Prueba de Wilcoxon	81
Anexo 2: Diseño experimental	82
Posible órdenes de las muestras.....	82
Asignación de las muestras	84
Anexo 3: Implementación de las técnicas estadísticas	87
Prueba binomial	87
Prueba DFC.....	87
Prueba chi cuadrado: “A” - “No A”	88
Prueba chi cuadrado: “Igual” - “Diferente”	89
Anexo 4: Simulación de los datos y ejecución de las pruebas	90
Prueba binomial: datos prueba triangular	90
Prueba DFC.....	90
Prueba “A no A”: Diseño monádico	91
Prueba “A no A”: Diseño pareado.....	91

Prefacio

La calidad de los productos ha sido tema de gran preocupación desde la época de la edad media, pero con la evolución de las técnicas de producción y la gran competitividad en el mercado, la se ha comenzado a asociar mayormente a su capacidad de adaptarse a los requerimientos y gustos del consumidor.

En toda industria de manufactura, generalmente se asegura la calidad mediante controles de procesos basados en pruebas físicas y químicas, sin embargo, para los productos en la industria alimenticia, farmacéutica, automotriz y de cosméticos, entran en juego ciertas características organolépticas como el sabor, olor y textura, para las cuales se necesita evaluar la percepción del ser humano para tomar decisiones oportunamente. De esta necesidad, surge el análisis sensorial, definido de acuerdo con la norma UNE-EN ISO 5492:2010 (Asociación Española de Normalización y Certificación, 2010) como: “ciencia relacionada con la evaluación de los atributos organolépticos de un producto mediante los sentidos”.

TasteLab es la primera empresa experta en análisis sensorial en España, constituida en 2016 como “*spin-off*” de la Universidad de Santiago de Compostela, a partir de un grupo de investigación con más de 30 años de experiencia en el área de análisis sensorial en colaboración con el departamento de Química Analítica, Nutrición y Bromatología y el departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela. TasteLab desarrolló el software SENSESBIT, un híbrido dentro de una plataforma “SaaS” (del inglés “Software as a Service”) que permite realizar de manera sencilla y ágil el proceso de análisis estadístico aplicado a datos sensoriales mediante el lenguaje de programación R (R Core Team, 2022).

Este trabajo fin de máster se centrará en los métodos estadísticos del análisis sensorial discriminativo, utilizado para analizar los datos sensoriales e identificar si existe alguna diferencia entre dos productos basado en la percepción de los jueces.

Actualmente, el software SENSESBIT es capaz de realizar la prueba discriminativa triangular de diferencias, por lo que este trabajo propone la implementación de pruebas discriminativas adicionales y el desarrollo de un diagrama para facilitar la selección de la prueba discriminativa más conveniente basado en la revisión sistemática de la literatura correspondiente a los métodos discriminantes.

Este trabajo está organizado en siete capítulos, desarrollados de la siguiente manera:

En el capítulo 1, se provee una breve introducción al análisis sensorial y las normas correspondientes a su aplicación.

En el capítulo 2 se presentará la revisión literaria sobre las distintas pruebas discriminativas y se realizará la recomendación de cada prueba de acuerdo con los factores limitantes del tipo de producto.

En el capítulo 3 se explicarán las bases teóricas de las técnicas estadísticas correspondientes a las pruebas presentadas para permitir el desarrollo de los próximos capítulos de este trabajo fin de máster.

En el capítulo 4 se presentará el análisis de la potencia estadística y tamaño muestral requerido basado en las pruebas presentadas en el capítulo 3.

En el capítulo 6 se desarrolla el diseño de experimento para la correcta recolección de los datos, de manera que se facilite la integración con el código a desarrollarse en el capítulo 6 para la implementación de las pruebas discriminativas adicionales en R.

Finalmente, en el capítulo 7 se simulan datos para ejemplificar el uso de las funciones desarrolladas en este trabajo fin de máster.

1 Análisis sensorial

1.1 Introducción

El objetivo principal del análisis sensorial es entender la relación entre los consumidores y los productos, por tanto, sus funciones básicas consisten en proveer medidas de datos sensoriales y realizar pruebas de hipótesis estadísticas para determinar diferencias y similitudes entre productos (Bi, 2015).

El análisis sensorial se ha definido como un método científico utilizado para evocar, medir, analizar e interpretar las respuestas a productos en la manera en que son percibidos mediante los cinco sentidos del ser humano (Stone & Sidel, 2004). Esta definición no difiere con la expuesta en la norma UNE-EN ISO 5492:2010, pero ha sido aceptada por varias organizaciones profesionales dentro de la industria, tales como el Instituto de Tecnólogos de Alimentos y la Sociedad Americana para Ensayos y Materiales; de hecho, esta definición nos permite simplificar y resumir las actividades en las cuales se basan los principios y prácticas del análisis sensorial (Bi, 2015):

- Evocar:

El análisis sensorial provee las indicaciones para minimizar los factores que pueden sesgar a los jueces durante las sesiones de evaluación de manera que puedan enfocarse en la experiencia sensorial. Por ejemplo, utilizar espacios individuales para cada juez y etiquetar las muestras de manera aleatoria.

- Medir:

A pesar de que el análisis sensorial estudia la percepción de los consumidores y por ende puede, en principio, interpretarse erróneamente como un estudio subjetivo, realmente es una ciencia cuantitativa en la cual se recolectan datos numéricos para establecer las relaciones entre las características de los productos y la percepción del ser humano. Por ejemplo, se evalúa la proporción de veces que los jueces discriminan cambios en los productos y se generan respuestas numéricas sobre la percepción.

- Analizar

Para evaluar si la relación observada entre las características de los productos y las respuestas sensoriales no son el resultado de variaciones no controlables durante el proceso de evaluación, tales como la motivación del participante o su sensibilidad a cierta estímulo sensorial, se utilizan métodos estadísticos para analizar los datos.

- Interpretar los resultados

La información estadística realmente sólo es útil cuando es interpretada en el contexto de hipótesis y con el conocimiento de las implicaciones de las decisiones a ser tomadas. Estas conclusiones implican la consideración del método utilizado, las limitaciones del experimento y el marco contextual del estudio.

1.2 Métodos de prueba de análisis sensorial

Las pruebas de análisis sensorial tienden a clasificarse de acuerdo con el objetivo del experimento, es decir, la pregunta que se busque responder sobre el producto a probar (Lawless & Heymann, 2010):

Pruebas discriminativas

El objetivo principal de las pruebas discriminativas es responder si existe alguna diferencia perceptible entre dos productos o si pueden considerarse similares.

El análisis de los datos generalmente se basa en estadísticos de frecuencias y proporciones, en términos básicos, una proporción de respuestas correctas por encima de lo esperado por el azar, es considerado evidencia para la existencia de una diferencia perceptible entre productos. En el caso de las prueba de diferencias, la hipótesis nula es “los productos no son diferentes”, y para las pruebas de similitud, la hipótesis nula es “los productos son iguales”.

Este trabajo fin de máster se centrará en la revisión literaria e implementación de pruebas discriminativas de análisis sensorial, por lo que en el capítulo 2 se presenta un análisis más detallado sobre las bases teóricas de estas pruebas.

Pruebas descriptivas

Los análisis descriptivos pretenden cuantificar cómo difieren ciertas características sensoriales entre los productos de la prueba. Estas pruebas se consideran exhaustivas porque la descripción puede contener aspectos cualitativos y cuantitativos al comparar las características de ambos productos y definir su intensidad.

Considerando que ciertas características descriptivas pueden relacionarse directamente con características físicas que pueden ser medidas con instrumentos industriales calibrados, estas pruebas descriptivas permiten relacionar estas medidas con la información de aceptación de los consumidores mediante técnicas estadísticas como análisis de regresión y correlación.

Pruebas afectivas

Las pruebas afectivas también son conocidas como aceptación del consumidor o pruebas de preferencia porque tienen como objetivo cuantificar qué tanto un producto es aceptado por el consumidor o cuáles productos prefiere.

Estas pruebas generalmente son realizadas sobre la base de una escala balanceada en una categoría central neutral y con adverbios que puedan ser relacionados con intervalos que puedan ser utilizados en el análisis estadístico.

1.3 Normas de análisis sensorial

El análisis sensorial requiere de métodos y procedimientos estandarizados para obtener resultados que sean consistentes y confiables, por tanto, se han establecidos normas y estándares para guiar varios aspectos del análisis sensorial, tales como los protocolos de prueba, selección y entrenamiento de los asesores y el análisis de los datos sensoriales.

En este acápite se proporcionará una revisión general de ciertas normas identificadas como normas para consulta de todas las pruebas discriminativas a evaluarse en este trabajo fin de máster. Cualquier consideración adicional que se identifique en las normas específicas para cada prueba estarán indicadas en cada acápite del capítulo 2.

1.3.1 UNE-EN ISO 5492:2010: Análisis sensorial. Vocabulario

La norma UNE-EN ISO 5492:2010 (Asociación Española de Normalización y Certificación, 2010) es la versión española del estándar internacional ISO 5492:2008: “Sensory Analysis – Vocabulary”, norma que define los términos relativos al análisis sensorial, agrupándolos en:

1. Terminología general
2. Terminología relativa a los sentidos
3. Terminología relativa a los atributos organolépticos
4. Terminología relativa a métodos

Considerando que este trabajo fin de máster se realiza con especial interés en los aspectos estadísticos de las pruebas, se hará hincapié en la terminología general y la terminología relativa a los métodos.

Terminología general

Término	Concepto
Análisis sensorial	Ciencia relacionada con la evaluación de los atributos organolépticos de un producto mediante los sentidos.
Calidad	Conjunto de rasgos y características de un producto, proceso o servicio, que confiere su aptitud para satisfacer las necesidades explícitas o implícitas.
Capacidad de discriminación	Sensibilidad, agudeza, capacidad para percibir diferencias cualitativas y/o cuantitativas.
Discriminación	Acción de diferenciación cualitativa y/o cuantitativa entre dos o más estímulos.
Evaluator sensorial; juez	Toda persona que toma parte de un ensayo sensorial.
Muestra	Espécimen o alícuota del producto presentado para evaluación.
Muestra control	Muestra del producto sometida al ensayo que se elige como referencia respecto a la cual se comparan las demás muestras

Término	Concepto
Muestra de referencia	Estímulo o sustancia, a veces diferente del material en ensayo, cuidadosamente seleccionado para definir o ilustrar un atributo, o un nivel específico de un atributo determinado, con el cual se comparan todos los demás.
Panel sensorial; jurado	Grupo de evaluadores que participan en un ensayo sensorial.
Producto	Material, comestible o no, que se puede evaluar mediante análisis sensorial.
Punto de referencia	Valor seleccionado (de uno o varios atributos o de un producto) respecto al cual se evalúan las muestras.

Tabla 1.1 - Terminología general

Terminología relativa a los métodos

Término	Concepto
Método objetivo	Cualquier método en el que se minimizan los efectos de las opiniones personales.
Clasificación	Método en el que se presume un nivel de calidad inherente a la escala empleada, con la finalidad de clasificar productos dentro de grupos de calidad.
Ordenación	Método de clasificación en el que una serie de dos o más muestras, presentadas simultáneamente, se ordenan según la intensidad o el grado de un determinado atributo.
Estimación de la magnitud	Procedimiento que asigna valores a las intensidades de un atributo, de manera tal que las relaciones entre los valores asignados sean las mismas que las relaciones entre las magnitudes de las percepciones correspondientes.
Evaluación independiente	Evaluación de uno o más estímulos sin comparación directa.
Ensayo de discriminación	Cualquier método de ensayo que implica la comparación entre muestras para determinar si las diferencias son perceptibles.
Ensayo de comparación por parejas	Método en el que los estímulos se presentan en pares para ser comparados en base a algunos criterios definidos
Escala de respuesta	Forma (por ejemplo, numérica, verbal, o utilización de imágenes) mediante la cual un evaluador registra una respuesta cuantitativa.
Sesgo	Error sistemático que puede ser positivo o negativo.
Sesgo previsible	Desvío debido a ideas preconcebidas de los evaluadores

Tabla 1.2 - Terminología relativa a los métodos

1.3.2 UNE-ISO 6658:2019: Análisis sensorial. Metodología. Guía general

La norma UNE-EN ISO 6658:2019 (Asociación Española de Normalización y Certificación, 2019) es la versión española del estándar internacional ISO 6658:2017: “Sensory analysis. Methodology. General guidance” y sirve como introducción general a la metodología de análisis sensorial.

Para fines de este trabajo fin de máster, se revisarán los aspectos más importantes para la implementación de las pruebas discriminativas desde un punto de vista estadístico, por tanto, se discutirán los capítulos relacionados a los requisitos generales, métodos de prueba, análisis de resultados e informe de la prueba; excluyendo los detalles específicos a cada prueba, que serán discutidos en el capítulo 2.

Requisitos generales

Los requisitos generales descritos en la norma buscan controlar y estandarizar la influencia que tienen los estímulos sensoriales externos en la respuesta de los evaluadores; identificar factores como la variabilidad en la respuesta sensorial de los evaluadores; y, diseñar el experimento de manera apropiada para controlar los sesgos sistemáticos.

Para las pruebas cuyo objetivo es distinguir entre dos o más productos, se identifica la necesidad de diferenciar entre saber:

- 1) Si existe alguna diferencia,
- 2) la magnitud de diferencia,
- 3) la dirección (o naturaleza) de esa diferencia,
- 4) la influencia de esa diferencia,
- 5) si toda o sólo una parte de la población detecta la diferencia.

La elección de la prueba depende del objetivo de esta, de factores asociados al producto, a los jueces, al entorno de la prueba y al nivel de precisión analítica y confianza estadística de las conclusiones.

Métodos de las pruebas

La norma UNE-EN ISO 6658:2019 hace distinción entre tres métodos de prueba:

- 1) Pruebas de discriminación para evaluar la probabilidad de la diferencia o similitud entre productos.
- 2) Pruebas que utilizan escalas o categorías para estimar el orden o el tamaño de las diferencias.
- 3) Pruebas descriptivas para caracterizar atributos sensoriales específicos.

En este capítulo se revisarán las consideraciones generales del primer método, las pruebas discriminativas, para las que, en principio, se debe especificar si el objetivo es demostrar si existe una diferencia o demostrar si existe una similitud significativa entre los productos.

Las pruebas utilizadas comúnmente para determinar la probabilidad de la diferencia o similitud en las muestras son las siguientes:

- 1) Prueba de comparación por parejas
- 2) Prueba triangular
- 3) Prueba dúo-trío
- 4) Prueba dos de cinco
- 5) Prueba “A”-“No A”
- 6) Prueba de tétradas (tetrad)

Análisis de resultados

De manera general, el análisis de las pruebas de discriminación se basa en los resultados de las pruebas proporcionados por los jueces en cada categoría, por ejemplo, aquellos que escogen correctamente la muestra diferente, o aquellos que identifican la muestra A o la muestra B con un mayor nivel del atributo sensorial en evaluación.

En las pruebas discriminativas donde los jueces no pueden indicar “no hay diferencia” (técnica de elección forzada), se obtienen respuestas de jueces que de otra forma hubieran respondido “no hay diferencia” pero pueden ser correctos. Sin embargo, esta técnica puede contradecir a los jueces que verdaderamente no identifican ninguna diferencia.

Informe de la prueba

Para garantizar la claridad y contextualización de la prueba en el informe, se recomienda indicar los siguientes puntos:

- Objetivo de la prueba
- Información del producto
- Métodos de preparación de la prueba
- Metodología utilizada
- Fecha y hora
- Condiciones materiales para realizar la prueba
- Característica del grupo de jueces
- Resultados: interpretación, modelo estadístico utilizado, modalidades y condiciones de acceso a los datos brutos.
- Conclusiones
- Recomendaciones
- Nombre e información de contacto de la persona responsable del estudio.

2 Pruebas discriminativas

Las pruebas discriminativas tienen la intención de evaluar si dos productos se perciben como diferentes o similares y sus conclusiones generalmente se basan en la proporción de jueces capaz de discriminar esta diferencia durante la sesión de evaluación.

En este capítulo se evaluarán los distintos métodos de pruebas discriminativas comúnmente utilizadas en el análisis sensorial, incluyendo sus diferencias en términos de diseño, procesos cognitivos, análisis estadísticos, y limitaciones para la aplicación de cada prueba, con el propósito de presentar una guía para la selección de la prueba adecuada según el objetivo y los factores limitantes del estudio.

Las pruebas discriminativas a revisar en este trabajo pueden ser clasificadas de acuerdo con las reglas de decisión y estrategias cognitivas utilizadas (Bi, 2015):

	Naturaleza de la diferencia requerida (Métodos especificados)	Comparación de la distancia de la diferencia (Métodos no especificados)
Métodos de selección forzada	2-AFC 3-AFC Tetrad especificada Diferente del control	Dúo-Trío Triangular Tetrad no especificada
Métodos con sesgo de respuesta	"A" – "No A"	Igual – diferente

Tabla 2.1 – Clasificación de las pruebas discriminativas

Los métodos de selección forzada requieren que los panelistas seleccionen la muestra "diferente" aun cuando el panelista realmente no detecta ninguna diferencia ni puede identificar cuál es la muestra que tiene el mayor nivel de la característica sensorial en evaluación.

El sesgo en la respuesta es la predisposición psicológica que tiene el panel para favorecer uno de los lados del criterio de la naturaleza de la diferencia entre dos muestras (por ejemplo, "igual" o "diferente").

Se podrían distinguir dos tipos de instrucciones en los métodos de prueba discriminativas (Bi, 2015):

1. Métodos especificados: se requiere indicar la naturaleza de la diferencia, por ejemplo, "¿Cuál muestra es más dulce?" o "¿La muestra es A o no es A?"
2. Métodos no especificados: se compara la distancia de la diferencia, por ejemplo, "¿Cuál de las dos muestras es igual a la muestra de control?", "¿Cuál de las tres muestras es la diferente?"

A pesar de que en principio parecieran preguntas similares, se asocian estrategias cognitivas diferentes, es decir, la pregunta es procesada de manera diferente por los panelistas, y, por ende, se pueden obtener resultados diferentes en términos de porcentaje de respuestas correctas.

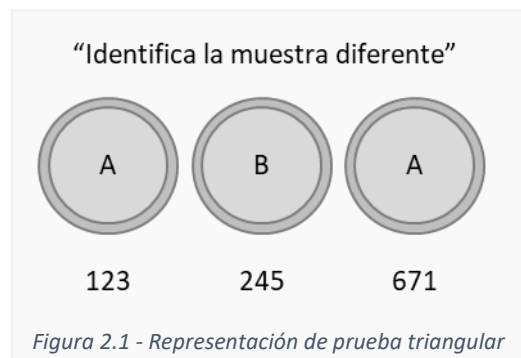
Considerando que la mayoría de estas pruebas comparten la técnica estadística utilizada en cada prueba, este capítulo se limitará a los aspectos generales de la prueba e indicar la técnica utilizada para luego presentarla en el capítulo 3.

2.1 Métodos de selección forzada

2.1.1 Prueba triangular

El principio de la prueba triangular consiste en presentar tres muestras, de las cuales dos se corresponden al mismo producto, esto con el propósito de que el juez identifique si existe alguna diferencia entre estas dos muestras y la tercera muestra presentada (Asociación Española de Normalización y Certificación, 2022).

En la prueba triangular, la segunda posición puede percibirse como la muestra diferente, aunque no lo sea, por lo que se utilizan seis posibles órdenes diferentes para contrarrestar este error psicológico: AAB, ABB, BAA, BBA, BAB y ABA (representado en la Figura 2.1); esta representación solo la conoce el experimentador, ya que al juez se le presentan códigos aleatorios para colocar como referencia en la respuesta.



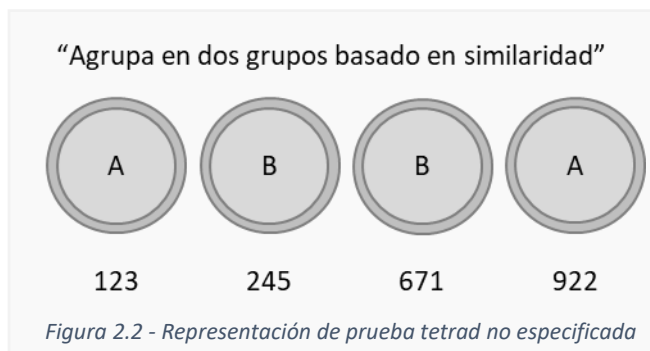
Al solo presentar tres muestras al juez, se obtiene una probabilidad de acertar al azar de $1/3$ ($p=0.333$) (Rogers, 2017). Dicho esto, la hipótesis nula de la prueba triangular establece que la probabilidad a largo plazo de realizar una decisión correcta cuando no hay una diferencia perceptible entre las muestras es de $1/3$; por el contrario, la hipótesis alternativa establece que la probabilidad de que la población subyacente tome la decisión correcta cuando perciben una diferencia entre las muestras es mayor a $1/3$ (Lawless & Heymann, 2010).

De acuerdo con lo expuesto en la Tabla 2.1, la prueba triangular se considera un método de selección forzada, por tanto, el juez no puede indicar que no percibe diferencia, sino que debe apuntar a la muestra que considera diferente, aun si esto implica seleccionar al azar.

El análisis e interpretación de los resultados se realiza basado en la cantidad mínima de respuestas correctas que son requeridas para concluir en una diferencia significativa de acuerdo con el nivel de significancia establecido. Esta cantidad mínima se puede encontrar en las tablas estadísticas de los estándares ISO 4120:2004 (Asociación Española de Normalización y Certificación, 2022) , sin embargo, considerando que su cálculo está basado en la distribución binomial, esto se discutirá a profundidad en el capítulo 3 al evaluar las técnicas estadísticas detrás de cada prueba de análisis sensorial.

2.1.2 Prueba tetrad

En la prueba tetrad, se presentan cuatro muestras: dos corresponden al producto A y dos corresponden al producto B. Se les solicita a los asesores que formen dos grupos basado en la similitud entre todas las muestras (tetrad no especificada) o que seleccionen las dos muestras con el mayor nivel de una característica sensorial (tetrad especificada) (Rogers, 2017).



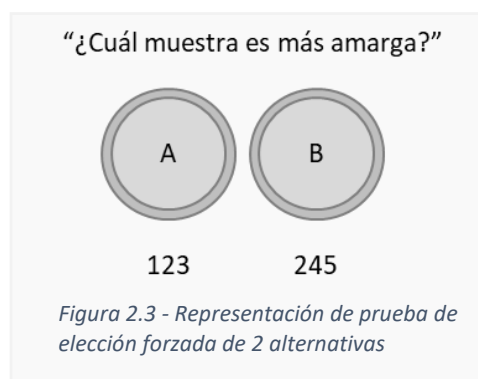
Al igual que con la prueba triangular, los resultados de esta prueba son analizados basado en la cantidad mínima de respuestas correctas requeridas para concluir que existe una diferencia significativa entre los grupos. De igual manera, su cálculo está basado en la distribución binomial con una probabilidad de acertar al azar de $1/3$ para la prueba tetrad no especificada, y una probabilidad de $1/6$ para la prueba tetrad especificada (American Society for Testing and Materials, 2016).

2.1.3 Prueba de elección forzada de 1 entre 2 alternativas (2-AFC)

La prueba de elección forzada de 2 alternativas se considera una prueba de comparación pareada direccionada ya que se comparan dos muestras sin describir la magnitud de la diferencia, pero determinan la dirección de la diferencia entre las dos muestras.

En estas pruebas se les instruye a los asesores para que indiquen cuál es la muestra que tiene el menor o mayor nivel de la característica sensorial o intensidad en general a evaluar.

La probabilidad de acertar al azar es de $\frac{1}{2}$ y los resultados se analizan mediante la prueba binomial (Asociación Española de Normalización y Certificación, 2007).



2.1.4 Prueba de elección forzada de 1 entre 3 alternativas (3-AFC)

La prueba de elección forzada de 3 alternativas (3-AFC) se asimila a la prueba triangular en términos de que se presentan 3 muestras, pero en vez de indicarse la diferencia en general entre las muestras, se indica cuál es la muestra que presenta el mayor o menor nivel, es decir, es una prueba direccionada.



El modelo estadístico asume que los datos se ajustan a una distribución binomial, por lo que, similar a la prueba triangular, los resultados de la 3-AFC se pueden analizar mediante la prueba binomial (Asociación Española de Normalización y Certificación, 2007).

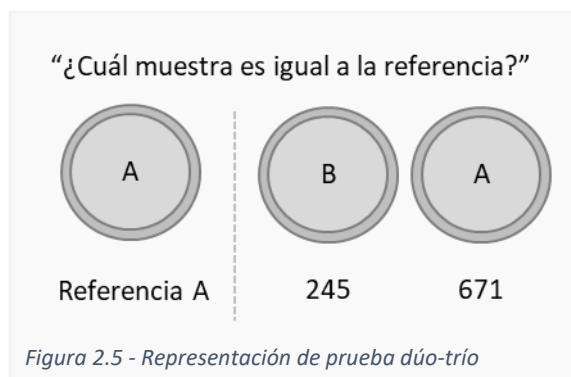
2.1.5 Prueba dúo-trío

La prueba dúo-trío se desarrolló como una alternativa a la prueba triangular y a las pruebas dual-estándar (Rogers, 2017). En esta prueba se le presentan tres muestras, una muestra de referencia y dos muestras adicionales para que indiquen cuál de las dos coincide con la muestra de referencia.

Al igual que la prueba “A” – “No A”, la prueba dúo-trío puede ser presentada en varias versiones (Rogers, 2017):

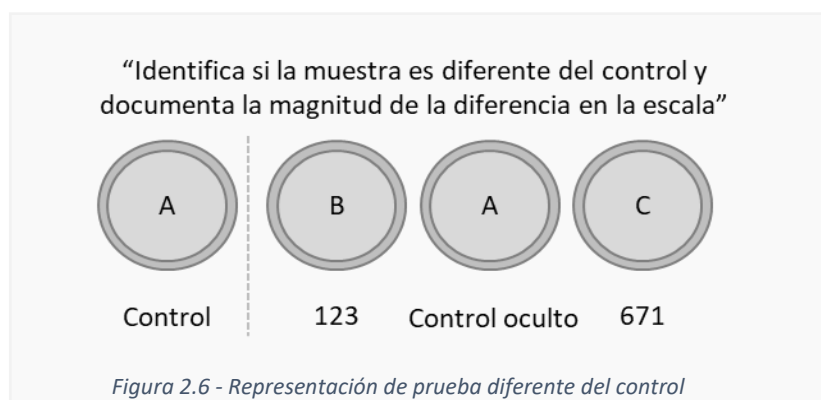
- Referencia constante: Siempre se presenta la referencia de la muestra que resulta menos familiar para los asesores. Es preferible para casos donde los asesores conocen mejor un producto que el otro.
- Referencia balanceada: Se utilizan ambas muestras como referencia, especialmente cuando ninguna de las muestras es conocida para los asesores.

La probabilidad de acierto al azar para estas pruebas es de $\frac{1}{2}$ y los resultados se analizan mediante la prueba binomial (Meilgaard, Civille, & Carr, 2016).



2.1.6 Prueba diferente del control (DFC)

En la prueba diferente del control, también conocida como prueba de grado de diferencia (Rogers, 2017), se presenta una muestra de control o referencia y se presentan muestras adicionales para que el juez evalúe qué tan diferente es cada muestra del control.



En adición, dentro de las muestras a evaluar, se presenta una muestra de control oculto para comparar los resultados de esta contra los resultados del resto de las muestras utilizando una escala de 0 - sin diferencia hasta 10-diferencia extrema, y se espera que la muestra de control esté dentro del rango de 0 a 2 para considerar válidos los resultados de la prueba (Kilcast, 2010). El resto de la escala se puede presentar como sigue (Muñoz, Civile, & Carr, 1992):

- 0 – No hay diferencia, igual que el control.
- 2 – Diferencia muy leve.
- 4 – Diferencia leve.
- 6 – Diferencia moderada.
- 8 – Diferencia grande.
- 10 – Diferencia extrema.

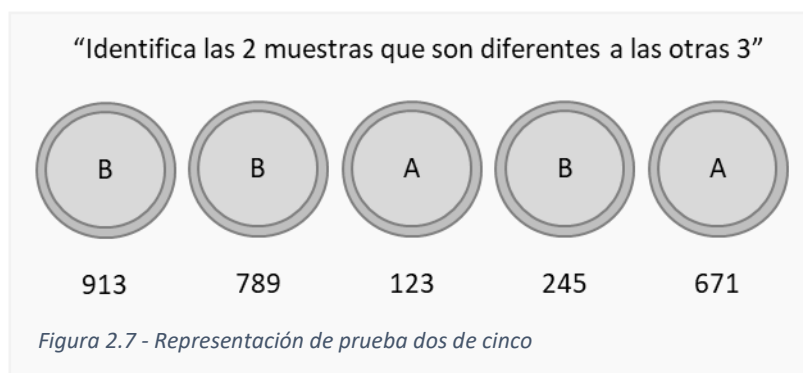
Los resultados de la prueba se analizan calculando la media de “diferente del control” para cada muestra y para el control oculto, luego se realiza una prueba T de Student si solo se está comparando una muestra contra el control, o un análisis de la varianza (ANOVA) para comparar dos o más muestras contra el control.

En caso de no cumplir con las asunciones requeridas para el uso de la prueba T de Student y ANOVA (a evaluarse en el capítulo 3), se recomienda utilizar las pruebas no paramétricas equivalentes: prueba de rango con signos de Wilcoxon y Prueba de Friedman, respectivamente.

2.1.7 Prueba dos de cinco

En la prueba dos de cinco, como su nombre puede ya indicarlo, al asesor se le presentan cinco muestras codificadas, de las cuales dos se corresponden a un producto y tres se corresponden a otro producto. La idea es que los asesores agrupen las muestras en un grupo de 2 y otro de tres.

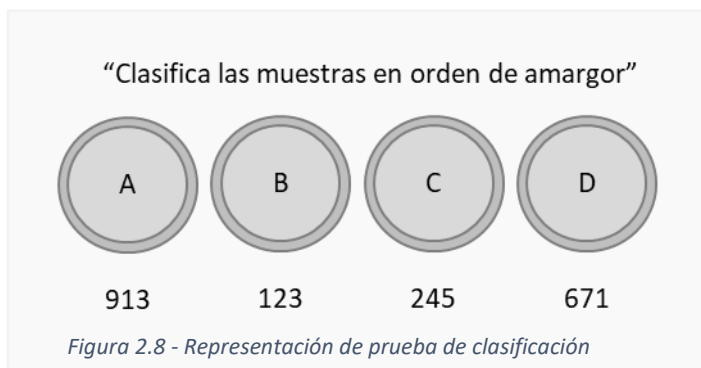
A pesar de que la prueba puede ser sensorialmente desgastante, es recomendable por ser estadísticamente muy eficiente al tener una probabilidad de acertar al azar de 1/10. Los resultados se pueden analizar mediante la prueba binomial (Meilgaard, Civile, & Carr, 2016)



2.1.8 Prueba de clasificación

En la prueba de clasificación, el asesor recibe muestras organizadas de manera aleatoria para que las clasifique en orden de acuerdo con el criterio a evaluar en la prueba. Generalmente, se solicita organizar las muestras desde la intensidad más débil a la más fuerte.

Considerando que el orden de las muestras en la escala no se corresponde con ninguna distancia o magnitud de la diferencia, se utiliza la técnica de suma de rango de Friedman cuando no se conoce el orden real de las muestras, y la prueba de Page cuando se puede establecer una hipótesis sobre el orden real de las muestras (Rogers, 2017).

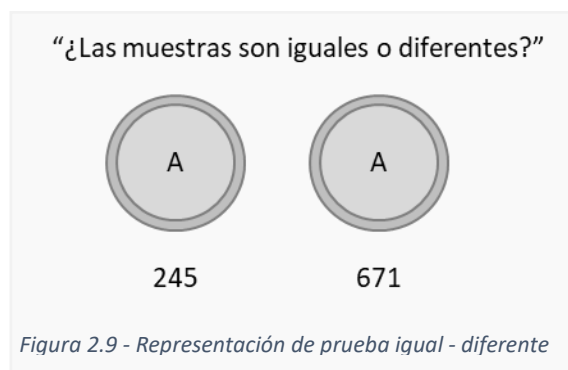


2.2 Métodos con sesgo de respuesta

2.2.1 Prueba igual – diferente

La prueba igual-diferente es un tipo de comparación pareada, también conocida como prueba de diferencia simple. Esta prueba busca determinar si existe una diferencia entre dos muestras, sin especificar la naturaleza de la diferencia.

Al asesor se le presentan dos muestras (Figura 2.9), solicitándole que indique si las muestras son iguales o diferentes. De la cantidad total de muestras presentadas, la mitad corresponden a pares diferentes, y la otra mitad corresponden a pares de la misma muestra presentada dos veces (Meilgaard, Civille, & Carr, 2016).



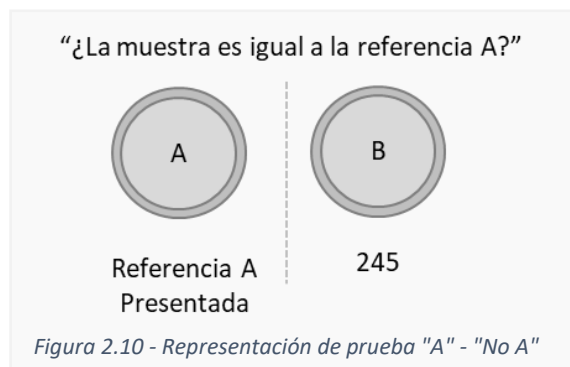
La prueba igual-diferente se puede realizar mediante un diseño monádico, mixto o pareado de acuerdo con las muestras evaluadas por los jueces. Los resultados se muestran en una tabla de contingencia y se evalúan mediante la prueba de chi cuadrado χ^2 de Pearson para homogeneidad, chi cuadrado χ^2 de Pearson para independencia o la prueba de McNemar para proporción correlacionada de acuerdo con el diseño utilizado (Bi, Sensory Discrimination Tests and Measurements, 2015).

2.2.2 Prueba “A” – “No A”

Esta prueba se define como un método de clasificación con dos categorías (Rogers, 2017), donde se requiere que el asesor identifique si la muestra presentada es “A” o “No A”.

En el caso presentado en la Figura 2.10, se le presenta la referencia “A” al asesor antes de que inicie la evaluación de la muestra (versión 2 de la Tabla 2.2). Este diseño tiene la ventaja de que los asesores evalúan las muestras en un contexto similar con respecto a la muestra de referencia; En el caso contrario, los

asesores tendrían que responder en base a la memoria de un entrenamiento previo y los resultados variarían en base a la experiencia del asesor. En la Tabla 2.2 se presentan las distintas versiones de presentación de muestras en la prueba “A” – “No A” (Rogers, 2017).



Prueba “A” – “No A”	Descripción	Ventajas	Desventajas
Versión 1	No se presenta ninguna prueba de referencia antes de la sesión de evaluación.	La prueba es realizada de manera rápida al confiar en la referencia interna de los asesores.	Se requieren asesores que conozcan el producto de referencia.
Versión 2	Se presenta la muestra de referencia “A” antes de la sesión de evaluación.	Los asesores son más capaces de detectar diferencias y están alineados en el concepto de la muestra de referencia “A”	Los evaluadores no saben qué esperar en términos de diferencias entre el producto “A” y el producto “No A”
Versión 3	Se presentan ambas muestras de referencia “A” y “No A” antes de la sesión de evaluación.	Los asesores están completamente alineados.	Se requiere más tiempo de entrenamiento, pero los asesores son más conscientes de sus respuestas.
Versión 4	Se presentan ambas muestras de referencia “A” y “No A” antes y durante la sesión de evaluación.	Conveniente en caso de que los asesores no conozcan el producto.	Es la versión que requiere más tiempo durante la sesión y no se recomienda para productos que tengan grandes efectos de arrastre.

Tabla 2.2 - Versiones de la prueba “A” - “No A”

Al igual que la prueba igual – diferente, los resultados de la prueba “A” - “No A” se presentan en una tabla de contingencia y pueden ser analizados mediante la prueba de chi cuadrado χ^2 de Pearson para homogeneidad, chi cuadrado χ^2 de Pearson para independencia o la prueba de McNemar para proporción correlacionada de acuerdo con el diseño utilizado (Bi, Sensory Discrimination Tests and Measurements, 2015).

2.3 Comparación y selección de las pruebas discriminativas

En este acápite, se presenta una revisión literaria para definir diferencias y consideraciones para la selección del tipo de prueba discriminativa a utilizar de acuerdo con ciertos factores, tales como las características del producto y asesores a utilizar.

Durante la sesión de evaluación, los asesores responden a una pregunta específica de acuerdo con la prueba seleccionada, la cual conlleva procesos cognitivos diferentes, por lo que, ciertas pruebas pueden ser más eficientes que otras y obtener más respuestas correctas a pesar de compartir características como la probabilidad de responder correctamente al azar.

Antes de presentar los detalles de ventajas y desventajas de cada prueba, una guía general para la elección de la prueba adecuada debe iniciar planteando el objetivo del proyecto para identificar si está alineado con el objetivo de las pruebas discriminativas. Debajo se presentan algunos ejemplos de tipos de problema y escenarios donde el analista se puede beneficiar del uso de pruebas discriminativas (Meilgaard, Civille, & Carr, 2016):

1. Identificar si existe diferencia entre un producto existente y un producto en desarrollo.
2. Confirmar si un producto se percibe diferente luego de cambiar el proceso de manufactura o composición química.
3. Determinar la dirección del cambio necesario para mejorar un producto.
4. Aseguramiento de la calidad del producto durante todas las etapas de la producción, manteniendo las mismas características del estándar.

Una vez que el objetivo está definido, se pueden considerar las áreas de aplicación, ventajas y desventajas de cada prueba:

Tipo de prueba	Prueba discriminativa	Consideraciones
Diferencia general	Prueba triangular	No es recomendada para muestras con efectos sobre la fatiga sensorial y/o arrastre de la característica sensorial (Meilgaard, Civille, & Carr, 2016) y en general, no provee información sobre la magnitud de la diferencia.
	Prueba tetrad	Al igual que la prueba triangular, posee limitaciones por la fatiga sensorial, pero una ventaja práctica de la prueba tetrad es que se requiere un menor número de panelistas que la prueba triangular, y detectaría diferencias menores de acuerdo con el modelo thurstoniano (Stone & Sidel, 2004).
	Prueba “Igual” – “Diferente”	La prueba “Igual”–“Diferente” es adecuada para muestras que producirían cierta fatiga sensorial (Meilgaard, Civille, & Carr, 2016).
	Prueba “A” – “No A”	Posee ventajas sobre las demás porque las muestras se presentan una a la vez, por lo que es útil para productos cuyas características

Tipo de prueba	Prueba discriminativa	Consideraciones
		sensoriales pueden cambiar durante la sesión de evaluación (Rogers, 2017).
	Prueba duo-trío	El uso de una muestra de referencia y un control oculto como estabilizador del criterio, hace que esta prueba sea conveniente para muestras con una característica sensorial intensa, siendo más relevante para atributos de sabor (Rogers, 2017).
	Prueba diferente del control (DFC)	La prueba DFC permite cuantificar el grado de diferencia entre los productos. En adición, la prueba DFC es útil para evaluar muestras no homogéneas, con sabor/olor fuerte y gran nivel de arrastre ya que la muestra control oculta permite obtener una base para la comparación (Rogers, 2017).
	Prueba dos de cinco	Esta prueba requiere una cantidad considerable de presentaciones hacia al asesor, por lo que se recomienda para situaciones donde sean prioridad las características de apariencia sobre sabor u olor (Rogers, 2017).
Diferencia de atributo	2-AFC	Se recomienda para productos con sabor fuerte o efecto de arrastre y cuando la variación lote a lote es tan grande como la diferencia entre los dos productos de la evaluación (Lawless & Heymann, 2010).
	3-AFC	La prueba 3-AFC posee las mismas limitaciones que la prueba triangular en términos de fatiga sensorial, su selección va depender de si el estudio busca evaluar diferencia por atributo de las muestras.
	Prueba de clasificación	Esta prueba no depende de la memoria del asesor ya que las muestra son presentadas simultáneamente y el asesor las organiza de acuerdo con el nivel de la característica sensorial. Sin embargo, se requiere verificarlos todas las muestras antes clasificarlas, por lo que aunque se recomienda para comparar muchas muestras en poco tiempo, no se recomienda para productos con efectos de arrastre (Stone & Sidel, 2004).

Tabla 2.3 - Consideraciones de las pruebas discriminativas

Basado en las consideraciones expuestas anteriormente, se propone el siguiente diagrama para la selección de la prueba discriminativa:

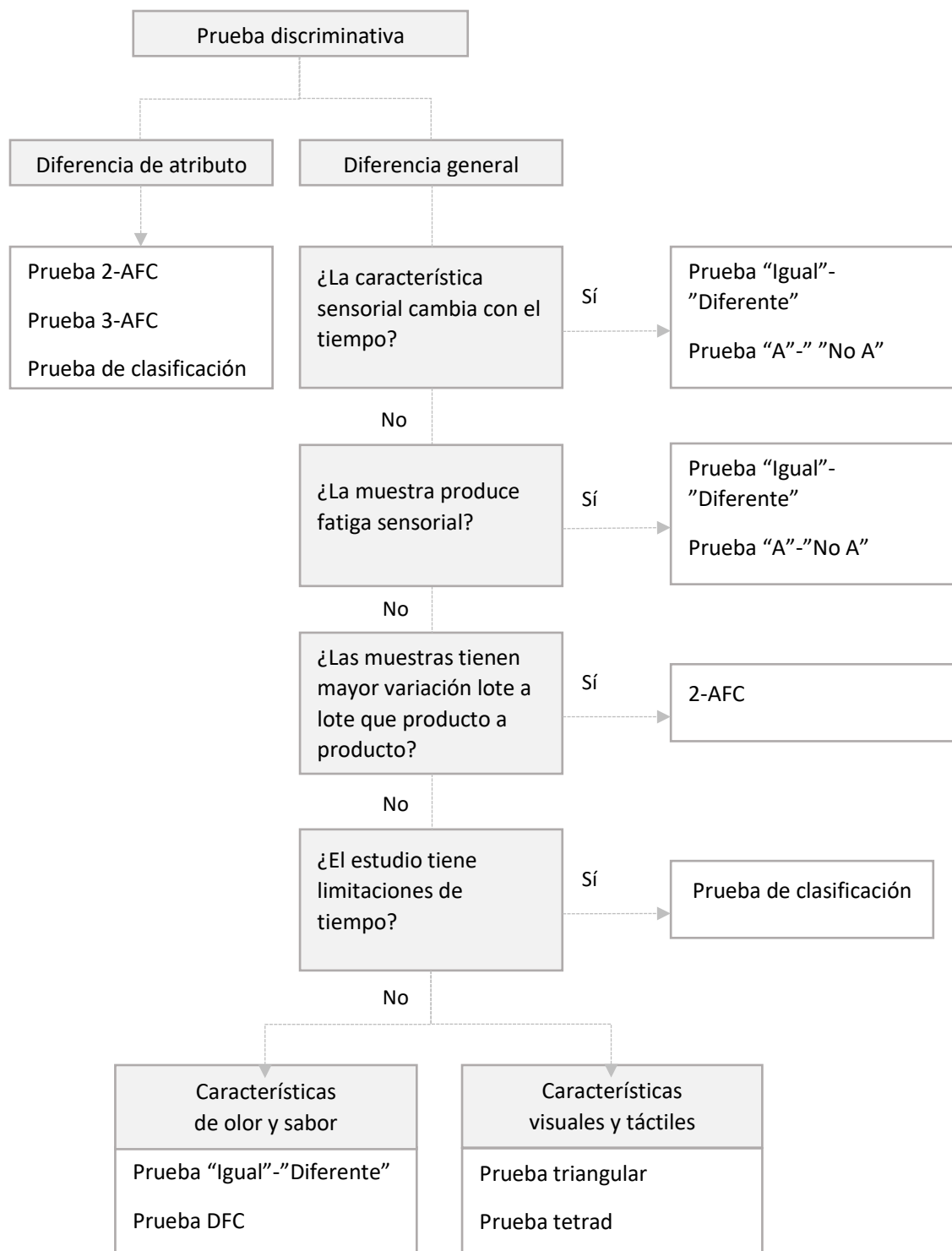


Figura 2.11 - Diagrama para selección de prueba discriminativa

Estas pruebas tienen diferentes niveles de potencia estadística, lo que refleja su capacidad para detectar una diferencia real entre muestras cuando realmente existe una, por lo que este aspecto también debe ser tomado en consideración para los escenarios donde la cantidad de jueces disponibles sea una limitante. Referirse al capítulo 4 de este trabajo fin de máster.

3 Técnicas estadísticas para pruebas discriminativas de diferencia

A modo de resumen, en la Tabla 3.1 se muestran las pruebas discriminativas asociadas a las técnicas estadísticas que se revisarán en este capítulo.

Método de prueba	Prueba discriminativa	Técnica estadística
Métodos de selección forzada	Prueba triangular	
	Prueba tetrad no especificada	
	Prueba tetrad especificada	
	Prueba dos de cinco	Binomial
	Prueba 2-AFC	
	Prueba 3-AFC	
	Prueba dúo-trío	
Métodos con sesgo de respuesta		Pruebas paramétricas: Prueba T de Student ó Análisis de la varianza (ANOVA)
	Prueba diferente del control (DFC)	Pruebas no paramétricas: Prueba de los rangos con signo de Wilcoxon ó Prueba de Friedman
	Prueba “A” – “No A” con diseño monádico.	χ^2 de Pearson para homogeneidad
	Prueba “Igual – Diferente” con diseño monádico.	
	Prueba “A” – “No A” con diseño mixto.	χ^2 de Pearson para independencia
	Prueba “Igual – Diferente” con diseño mixto.	
	Prueba “A” – “No A” con diseño pareado.	χ^2 de McNemar para proporción correlacionada
	Prueba “Igual – Diferente” con diseño pareado.	

Tabla 3.1 - Técnicas estadísticas de las pruebas discriminativas

La prueba de clasificación se presenta en el capítulo 2 con el propósito de evaluar su uso para el diagrama mostrado en la Figura 2.11, sin embargo, esta prueba está implementada en el software SENSEBIT mediante el uso de la prueba de suma de rangos de Friedman, por lo que, los esfuerzos de este trabajo se centrarán en la integración de las pruebas descritas en la Tabla 3.1.

3.1 Prueba binomial

Una distribución binomial resulta de un experimento que posee las siguientes características:

- Número fijo de ensayos, n .
- Los ensayos son independientes.
- Los resultados de los ensayos se clasifican en dos categorías: éxito y fracaso.
- La probabilidad de éxito p , y probabilidad de fracaso q , se mantienen constante durante todos los ensayos.

El número de respuestas correctas en una prueba discriminativa se asume que es una variable binomial que sigue una distribución binomial $X \sim Bin(n, p)$. La probabilidad de observar x éxitos en n ensayos independientes de un experimento binomial sigue la siguiente función de masa de probabilidad (PDP) (Bi, 2015):

$$\Pr(X = x; p, n) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, 2, \dots, n \quad (3.1.1)$$

Siguiendo la misma notación, la función de distribución acumulativa (CDF) viene dada por:

$$F(x) = \sum_{k=0}^x \binom{n}{k} p^k (1 - p)^{n-k} \quad (3.1.2)$$

En el contexto del análisis sensorial, la prueba binomial se utiliza como contraste de hipótesis para evaluar si el número de respuestas correctas en la prueba discriminativa es significativamente diferente a los resultados que se esperarían por casualidad.

Jian Bi presenta las pruebas de diferencias basada en un modelo de adivinanza donde se indica la relación entre tres cantidades:

- p_c : probabilidad de respuestas correctas
- p_0 : probabilidad de respuestas correcta por adivinanza
- p_d : probabilidad de discriminar

La relación es como sigue (Bi, 2015):

$$p_c = p_d + p_0 (1 - p_d) \quad (3.1.3)$$

Si los productos son iguales, un discriminador provee una respuesta correcta con una probabilidad de 1 y un no discriminador provee una respuesta correcta con una probabilidad por adivinanza, p_0 . Por tanto, la probabilidad de que cada asesor provea una respuesta correcta viene dada por la ecuación (3.1.3).

Considerando la notación de este modelo, los pasos para realizar la prueba binomial serían como sigue (Mendenhall, Beaver, & Beaver, 2013) y (Bi, 2015):

1. Definir la hipótesis nula y alternativa

La hipótesis nula indica que no hay diferencia entre las muestras:

$$H_0: p_c = p_0 \quad (3.1.4)$$

La probabilidad de respuestas correctas por adivinanza, p_0 , es un valor fijo que depende de las posibles alternativas de selección de acuerdo con la prueba discriminativa utilizada:

Prueba discriminativa	Probabilidad de adivinar (p_0)
Prueba duo – trio	1/2
Prueba 2-AFC	1/2
Prueba 3-AFC	1/3
Prueba triangular	1/3
Prueba tetrad no especificada	1/3
Prueba tetrad especificada	1/6
Prueba dos de cinco	1/10

Tabla 3.2 - Probabilidades de adivinanza de las pruebas discriminativas

En general, la hipótesis alternativa indicaría que hay una diferencia entre las muestras comparadas, pero se debe definir si la prueba es unilateral (“one-sided test”) o bilateral (“two-sided test”) dependiendo del propósito de la evaluación:

- **Prueba unilateral:**

Se utiliza cuando se tiene una hipótesis clara sobre la dirección de la diferencia entre las dos muestras, por ejemplo, cuando se espera que una de las muestras tenga un mayor nivel de la característica sensorial de interés. La formulación matemática sería como sigue:

$$H_1: p_c > p_0 \quad (3.1.5)$$

- **Prueba bilateral:**

En este caso, se busca detectar si existe una diferencia entre las muestras sin indicar si alguna de las muestras es mayor o menor en términos de la característica sensorial. La formulación matemática sería como sigue:

$$H_1: p_c \neq p_0 \quad (3.1.6)$$

A pesar de que varios autores recomiendan que las pruebas bilaterales sean utilizadas siempre que la hipótesis de investigación sea no direccional, hay que evaluar la relación entre esta hipótesis y la hipótesis estadística (Cho & Abe, 2012).

Considerando que mediante la prueba binomial se está evaluando si la proporción de éxitos es mayor a lo que se esperaría por el azar y al evaluar las instrucciones que se proveen al asesor para llevar a cabo la prueba (ver Tabla 3.3), se puede recomendar el uso de la prueba estadística unilateral aunque la hipótesis de investigación sea bilateral (que las muestras son diferentes), ya que si el asesor identifica una diferencia general, se espera que seleccione la muestra que realmente es diferente del resto; Por lo que, realizar una prueba bilateral implicaría la posibilidad de rechazar la hipótesis nula cuando la proporción de asesores que identifican la muestra diferente sea considerablemente menor a lo esperado al seleccionar al azar (cola izquierda), lo que realmente podría significar que las muestras no poseen una diferencia perceptible y no debería plantear duda sobre la hipótesis nula.

Para las pruebas direccionadas o especificadas (prueba 2-AFC, prueba 3-AFC, y prueba tetrad-especificada), puede ser conveniente utilizar la prueba bilateral cuando no se tiene la certeza de cuál de las dos muestras realmente tiene un mayor nivel de la característica a evaluar; en este caso, una proporción significativamente por debajo de lo que se esperaría al azar (cola izquierda) podría implicar que existe una diferencia contraria a lo que se estableció en la hipótesis de investigación.

Prueba discriminativa	Instrucción al asesor	Prueba	Racional
Prueba duo – trio	“¿Cuál muestra es igual a la referencia?”	Hipótesis de investigación: Las muestras son diferentes Prueba estadística: Unilateral hacia la derecha	Si el asesor puede distinguir entre las dos muestras, se espera que seleccione la muestra que coincide con la referencia. Una proporción estadísticamente menor que el azar, no debería llevar a rechazar la hipótesis nula porque implica que no hay diferencia perceptible.
Prueba 2-AFC	“Cuál muestra tiene el mayor nivel de la característica sensorial”	Hipótesis de investigación: La muestra A posee un mayor nivel que la muestra B Prueba estadística: Unilateral hacia la derecha o <u>bilateral</u>	Se puede realizar una prueba bilateral si no se tiene certeza de cuál muestra realmente es mayor en la característica a evaluar. Si la proporción es estadísticamente menor a lo esperado al azar, la diferencia es en dirección contraria a lo expuesto en la hipótesis de investigación.
Prueba 3-AFC			

Prueba discriminativa	Instrucción al asesor	Prueba	Racional
Prueba triangular	“Identifica la muestra diferente”	<p>Hipótesis de investigación:</p> <p>Las muestras son diferentes</p> <p>Prueba estadística: Unilateral hacia la derecha</p>	<p>Si el asesor puede distinguir entre las tres muestras, se espera que seleccione la muestra que es diferente.</p> <p>Una proporción estadísticamente menor que el azar, no debería llevar a rechazar la hipótesis nula porque sugiere que no hay diferencia perceptible.</p>
Prueba tetrad no especificada	“Agrupa en dos grupos basado en similaridad”	<p>Hipótesis de investigación:</p> <p>Las muestras son diferentes</p> <p>Prueba estadística: Unilateral hacia la derecha</p>	<p>Si el asesor puede distinguir entre los dos grupos, se espera que agrupe correctamente.</p> <p>Una proporción estadísticamente menor que el azar, no debería llevar a rechazar la hipótesis nula porque sugiere que no hay diferencia perceptible.</p>
Prueba tetrad especificada	“Selecciona las dos muestras con el mayor nivel de la característica sensorial”	<p>Hipótesis de investigación:</p> <p>La muestra A posee un mayor nivel que la muestra B</p> <p>Prueba estadística: Unilateral hacia la derecha o <u>bilateral</u></p>	<p>Se puede realizar una prueba bilateral si no se tiene certeza de cuál muestra realmente es mayor en la característica a evaluar.</p> <p>Si la proporción es estadísticamente menor a lo esperado al azar, la diferencia es en dirección contraria a lo expuesto en la hipótesis de investigación.</p>
Prueba dos de cinco	“Identifica las 2 muestras que son diferentes a las otras 3”	<p>Hipótesis de investigación:</p> <p>Las muestras son diferentes</p> <p>Prueba estadística: Unilateral hacia la derecha</p>	<p>Si el asesor puede distinguir entre los dos grupos, se espera que agrupe correctamente.</p> <p>Una proporción estadísticamente menor que el azar, no debería llevar a rechazar la hipótesis nula porque sugiere que no hay diferencia perceptible.</p>

Tabla 3.3 – Instrucciones de las pruebas discriminativas, hipótesis de investigación y prueba estadística

2. Determinar el nivel de significancia de la prueba, α .

El nivel de significancia controla la probabilidad de rechazar la hipótesis nula cuando realmente es cierta (error tipo I). Convencionalmente, se selecciona una probabilidad α de 0.1, 0.05 ó 0.01 (Bower, 2013).

3. Calcular el estadístico de la prueba

El estadístico de la prueba basado en la distribución binomial es el número de respuestas correctas, X .

4. Calcular el valor crítico y el p-valor de la prueba

▪ Valor crítico

El valor crítico de la prueba representa el número de respuestas correctas requeridas para rechazar la hipótesis nula basado en el nivel de significancia seleccionado (Bi, 2015).

El valor crítico para la prueba unilateral es:

$$\sum_{k=0}^c \binom{n}{k} p^k (1-p)^{n-k} \geq 1 - \alpha \quad (3.1.7)$$

El valor crítico para la prueba bilateral es:

$$\sum_{k=0}^c \binom{n}{k} p^k (1-p)^{n-k} \geq 1 - \alpha/2 \quad (3.1.8)$$

Las ecuaciones (3.1.7) y (3.1.8) representan la función inversa de la función de distribución acumulativa (CDF), también conocida como la función cuantil (Blitzstein & Hwang, 2015). Esta función busca el valor de la variable en el cual se obtiene un valor de probabilidad deseada, en otros términos, busca el valor crítico de la prueba.

Los cálculos se podrían realizar manualmente o utilizando las tablas de valores críticos para predeterminados valores de tamaño muestral y nivel de significancia que proveen los libros de texto de estadística y organizaciones como NIST (“National Institute of Standards and Technology”). En adición, se puede utilizar la función ‘qbinom’ del paquete base ‘stats’ de R (R Core Team, 2022), utilizando los siguientes argumentos:

Argumento	Descripción
p	El vector de probabilidades representa el nivel de significancia α seleccionado. En el caso de la prueba de dos colas, se colocaría el nivel de significancia dividido entre dos para evaluar la significancia estadística en ambas direcciones.
size	Cantidad de ensayos.
prob	Probabilidad de respuestas correctas en cada ensayo. Se selecciona la probabilidad indicada en la Tabla 3.2 de acuerdo con la prueba discriminativa realizada.
lower.tail	Este argumento está indicado como “TRUE” por defecto, lo que calcularía el valor crítico para la cola inferior de los valores. Al indicar “FALSE” se

Argumento	Descripción
	<p>calcularía el valor crítico para la cola superior, por lo que esta sería la opción adecuada para la hipótesis alternativa unilateral $H_1: p_c > p_0$.</p> <p>Para la prueba alternativa bilateral $H_1: p_c \neq p_0$, se necesitaría evaluar el valor crítico en ambas colas, por lo que se haría un cálculo con el argumento lower.tail en "TRUE" y otro cálculo para el argumento lower.tail en "FALSE", considerando dividir el nivel de significancia α entre dos.</p>

Tabla 3.4 – Argumentos de función "qbinom" en R

▪ P-valor

El p-valor representa el menor nivel de significancia para el cual los datos observados indican que se debería rechazar la hipótesis nula basado en el estadístico observado (Wackerly, Mendenhall, & Scheaffer, 2008).

Al igual que el valor crítico, el p-valor puede ser calculado utilizando el paquete base 'stats' de R (R Core Team, 2022) mediante la función 'pbinom'. Esta función calcula la probabilidad acumulada de la distribución binomial hasta el valor introducido, en este caso, sería la probabilidad acumulada del estadístico calculado en la cantidad de ensayos realizados. Los argumentos de la función se completarían como sigue:

Argumento	Descripción
x	La cantidad de éxitos observados, es decir, la cantidad de evaluadores que identificaron la diferencia entre las muestras.
size	Cantidad de ensayos.
prob	Probabilidad de respuestas correctas en cada ensayo. Se selecciona la probabilidad indicada en la Tabla 3.2 de acuerdo con la prueba discriminativa realizada.
lower.tail	<p>Este argumento está indicado como "TRUE" por defecto, lo que calcularía la probabilidad acumulada de obtener una cantidad igual o menor de los éxitos indicados con el argumento 'x'. Al indicar "FALSE" se calcularía la probabilidad acumulada de obtener una cantidad igual o mayor.</p> <p>Para la hipótesis alternativa unilateral $H_1: p_c > p_0$, se utilizaría lower.tail="TRUE".</p> <p>Para la prueba alternativa bilateral $H_1: p_c \neq p_0$, se necesitaría sumar el resultado de evaluar la probabilidad en los dos extremos, por lo que se haría un cálculo con el argumento lower.tail en "TRUE" y otro cálculo para el argumento lower.tail en "FALSE".</p>

Tabla 3.5 - Argumentos de la función 'pbinom' en R

5. Conclusiones

Las conclusiones de la prueba binomial se pueden basar en dos enfoques:

- **Método de valor crítico**

El valor crítico separa la región de crítica, donde se rechaza la hipótesis nula, de los valores del estadístico de prueba donde no se rechaza la hipótesis nula (Triola, 2018). La región crítica depende de la formulación de la hipótesis alternativa:

Prueba bilateral: La región crítica se encuentra en las dos regiones extremas debajo de la curva, por lo que el valor del estadístico se compara con el valor crítico en ambas colas. Se rechaza la hipótesis nula si el estadístico es menor o igual que el valor crítico de la cola izquierda, o si es mayor o igual que el valor crítico de la cola derecha.

Prueba unilateral: La región crítica se encuentra en uno de los extremos, por lo que, se rechaza la hipótesis nula si el valor del estadístico es mayor o igual que el valor crítico de la prueba hacia la derecha.

- **Método de p-valor**

Siendo el p-valor la probabilidad de obtener un valor del estadístico de prueba que sea al menos tan extremo como el estadístico de prueba obtenido de la muestra, suponiendo que la hipótesis nula es cierta (Triola, 2018), la hipótesis nula se rechaza cuando se obtiene un p-valor menor al valor de significancia definido para la prueba.

En contrastes de hipótesis, ambos métodos son usados y su elección puede depender simplemente de los resultados que provee el software utilizado para el análisis estadístico. El valor crítico puede ser más intuitivo y fácil de interpretar; mientras que el p-valor provee información por sí solo sobre qué tan fuerte es la evidencia contra la hipótesis nula ya que no es necesario calcularlo para cada nivel de significancia.

Con el propósito de disminuir los cálculos, se realizará la prueba binomial mediante la función 'binom.test()' del paquete base 'stats' de R (R Core Team, 2022):

Argumento	Descripción
x	La cantidad de éxitos observados, es decir, la cantidad de evaluadores que identificaron la diferencia entre las muestras.
n	Cantidad de ensayos.
p	Probabilidad de respuestas correctas en cada ensayo. Se selecciona la probabilidad indicada en la Tabla 3.2 de acuerdo con la prueba discriminativa realizada.

Tabla 3.6 - Argumentos de la función 'binom.test()' en R

3.2 Prueba T de Student

La prueba T de Student es una prueba paramétrica designada para determinar si las medias de dos muestras son significativamente diferentes. En el contexto del análisis sensorial, esta prueba se utiliza en la prueba “Diferente del control (DFC)” para determinar si existe una diferencia entre la respuesta media para el control oculto y la muestra presentada durante la sesión de evaluación (O'Mahony, 1986).

Los tipos de prueba T de Student se pueden clasificar como sigue:

- **Prueba T de Student para dos muestras:**

Muestras pareadas o dependientes: evalúa la diferencia significativa entre las medias de dos muestras relacionadas, por ejemplo, los resultados dados por un mismo juez bajo condiciones distintas.

Muestras no pareadas o independientes: evalúa la diferencia significativa entre las medias de dos muestras no relacionadas, por ejemplo, los resultados dados por distintos jueces bajo condiciones distintas.

- **Prueba T de Student para una muestra:**

Evalúa si una muestra proviene de una población con media conocida.

Considerando que cada juez evalúa la muestra de control oculto contra el control y la muestra adicional contra el control durante la prueba DFC, el enfoque de este acápite será en la prueba T de Student para dos muestras pareadas.

El procedimiento para la prueba T de Student para dos muestras pareadas consiste en calcular la diferencia de las puntuaciones dadas por cada juez para la muestra control y la muestra a evaluar. Se espera que la media de las diferencias de estas muestras sea cercana a 0 si no existe una diferencia significativa entre los dos productos.

La hipótesis nula indica que la media de las diferencias de puntuaciones (\bar{d}) proviene de una población de puntuaciones con media igual a 0. Se expresaría como sigue:

$$H_0: \bar{d} = 0 \quad (3.2.1)$$

En contraste, la hipótesis alternativa indica que la media de las diferencias de puntuaciones proviene de una población de puntuaciones con media diferente de 0. Se expresaría como sigue:

$$H_1: \bar{d} \neq 0 \quad (3.2.2)$$

El estadístico de la prueba T de Student para muestras pareadas se calcula como sigue:

$$t: \frac{\bar{d}}{S/\sqrt{N}} \quad (3.2.3)$$

donde:

\bar{d} es la media de las diferencias de puntuaciones.

N es el tamaño muestral.

S es la desviación estándar de las diferencias de puntuaciones, calculada como sigue:

$$S = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{N}}{N-1}} \quad (3.2.4)$$

En la práctica, para el cálculo del estadístico y el p-valor de la prueba se utilizará la función 't.test()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
x	Vector conteniendo las puntuaciones de diferencias para la muestra control.
y	Vector conteniendo las puntuaciones de diferencias para la muestra en evaluación, estos valores deben estar pareados con los introducidos en el argumento 'x'.
alternative	Este argumento indica la hipótesis alternativa y debería indicarse como "two.sided" para calcular el p-valor considerando ambas colas, por lo que esta sería la opción adecuada para la hipótesis alternativa bilateral $H_1: \bar{d} \neq 0$.
mu	Argumento para indicar el valor de la media a comparar en la hipótesis, en este caso sería 0.
paired	Argumento para indicar si se realizará una prueba T de Student para muestras pareadas. Para la prueba DFC debe indicarse como "TRUE".

Tabla 3.7 - Argumentos de la función 't.test()' en R

Bajo la hipótesis nula, este estadístico sigue una distribución T de Student con n-1 grados de libertad, comúnmente conocida como "distribución T". En general, esta distribución tiene la misma forma de campana simétrica que la distribución normal estándar, pero con más variabilidad. Debajo se presentan algunas propiedades de esta distribución:

1. Al igual que la distribución normal estándar, es simétrica alrededor de la media de $t=0$.
2. La desviación estándar varía según el tamaño de la muestra, siendo siempre mayor que 1.
3. La distribución T de Student se aproxima a la distribución normal estándar conforme el tamaño muestral aumenta.

La función de densidad de densidad de probabilidad (PDF) es:

$$f(t) = \left\{ \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} \right\} \left(1 + \frac{t^2}{n} \right)^{-(n+1)/2}, -\infty < t < \infty \quad (3.2.5)$$

donde Γ es la función gamma y n es el tamaño muestral (Blitzstein & Hwang, 2015).

Tomando en cuenta las propiedades de la distribución T de Student, para utilizar la prueba T de Student de muestras pareadas, se deben tomar en consideración las siguientes asunciones de la prueba (O'Mahony, 1986):

1. Las diferencias de las puntuaciones provienen de una población con distribución normal, por lo que, los datos tienen que estar en escala de intervalos o ratio.

2. Las puntuaciones se muestrean aleatoriamente y son independientes entre sí.
3. Las diferencias de las puntuaciones se calculan entre las muestras evaluadas por un mismo juez (pareadas).

En la práctica, las asunciones de independencia y muestras pareadas se asumen de acuerdo con las condiciones de la recolección de los datos. Por otro lado, la asunción de normalidad se puede verificar mediante la prueba de Shapiro Wilk para normalidad con la función 'shapiro.test ()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
x	Vector conteniendo las diferencias de las puntuaciones.

Tabla 3.8 - Argumentos de la función 'shapiro.test()' en R

Para concluir, se rechazaría la hipótesis nula si el p-valor arrojado resulta ser más pequeño que el nivel de significancia seleccionado, entonces, se concluiría que sí existe diferencia significativa entre las muestras y el control.

3.3 Análisis de la varianza (ANOVA)

El modelo de análisis de la varianza es una prueba paramétrica que busca determinar si hay diferencia entre las medias poblacionales, pero permitiendo hacer esta comparación para más de dos grupos, por lo cual se puede considerar como una extensión de la prueba T de Student (Bower, 2013); Por tanto, el análisis de la varianza (ANOVA) se utiliza para analizar los resultados de la prueba “Diferente del control (DFC)” cuando se evalúan más de dos muestras.

La lógica detrás de la prueba de análisis de la varianza se basa en verificar la varianza de las puntuaciones dentro de un grupo o factor (debido a error experimental, como diferencias entre los jueces o variables experimentales que no son controlables) contra la varianza de las puntuaciones entre los grupos o factores. Este análisis se puede resumir como sigue (O'Mahony, 1986):

- Si la varianza entre grupos es mayor a la varianza dentro de los grupos, las medias son significativamente diferentes y el grupo tiene un efecto en la variable respuesta.
- Si la varianza entre grupos es aproximadamente igual a la varianza dentro de los grupos, las medias no son significativamente diferentes y el grupo no tiene efecto en la variable respuesta.

Para la prueba diferente del control (DFC) se realiza un análisis de la varianza (ANOVA) de un factor con diseño de bloque aleatorio (“randomized block design: repeated measures”), donde solo hay un factor que puede afectar la variable respuesta y se aísla el efecto que puede tener la variación en la respuesta debido a los distintos jueces (bloques) de la sesión de evaluación. En este diseño cada juez va a evaluar todas las muestras, por lo que, el número total de observaciones es $n = bk$, siendo b la cantidad de bloques (jueces) y k la cantidad de tratamientos (productos). (Mendenhall, Beaver, & Beaver, 2013).

La hipótesis nula indica que las medias de los tratamientos no difieren lo suficiente para concluir que las medias provienen de distintas distribuciones. Se expresaría como sigue:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad (3.3.1)$$

En contraste, la hipótesis alternativa indicaría que al menos una de las medias es diferente del resto.

H_A : Al menos una de las muestras es diferente del resto

El estadístico de la prueba (F ratio) se calcula como sigue:

$$F = \frac{MST}{MSE} \quad (3.3.2)$$

donde:

MST es la estimación de la varianza total entre los tratamientos y se calcula como sigue:

$$MST = \frac{SST}{k-1} = \frac{\sum \frac{T_i^2}{b} - \frac{x_{ij}^2}{n}}{k-1} \quad (3.3.3)$$

MSE es la estimación de la varianza total dentro de los tratamientos o por error experimental:

$$MSE = \frac{SSE}{(b-1)(k-1)} = \frac{Total\ SS - SST}{(b-1)(k-1)} = \frac{\left(\sum x_{ij}^2 - \frac{x_{ij}^2}{n}\right) - \left(\sum \frac{T_i^2}{b} - \frac{x_{ij}^2}{n}\right)}{(b-1)(k-1)} \quad (3.3.4)$$

En la práctica, para el cálculo del estadístico y el p-valor de la prueba se utilizará la función 'anova ()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
object	Objeto conteniendo los resultados del ajuste del modelo de análisis de la varianza

Tabla 3.9 - Argumentos de la función 'anova()' en R

El ajuste del modelo de análisis de la varianza se realiza mediante la función 'aov ()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
formula	<p>Argumento para introducir la fórmula del modelo. Para la prueba DFC, se podría colocar: Respuesta ~ Producto + Juez.</p> <p>donde:</p> <ul style="list-style-type: none"> ▪ 'Respuesta' es la variable que contiene la puntuación dada para cada muestra (control, muestra A, muestra B, etc) por cada juez. ▪ 'Producto' es la variable que contiene la identificación de la muestra (control, muestra A, muestra B, etc). ▪ 'Juez' es la variable que contiene el juez que proveyó la respuesta. <p>Esta fórmula indicaría el interés en comparar las medias de los diferentes productos (tratamientos) teniendo en cuenta la variabilidad entre los jueces (bloque).</p>
data	Dataframe conteniendo las variables especificadas en el argumento 'formula'.

Tabla 3.10 - Argumentos de la función 'aov()' en R

Bajo la hipótesis nula, este estadístico sigue una distribución F (también conocida como distribución Fisher-Snedecor) con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador. En general, la distribución F representa la distribución de la relación S_1^2/S_2^2 si se repite el proceso de selección de muestras de dos poblaciones distribuidas normalmente (Triola, 2018). Esta distribución posee las siguientes propiedades:

1. La distribución F no es simétrica
2. Los valores de la distribución F son no negativos.
3. La forma de la distribución F depende de los grados de libertad del numerador y denominador del estadístico.

Tomando en cuenta las propiedades de la distribución F, para realizar un análisis de la varianza (ANOVA), se deben tomar en consideración las siguientes asunciones de la prueba (Triola, 2018):

1. Las muestras son aleatorias e independientes entre sí.
2. Las puntuaciones provienen de una población con distribución normal.

Las asunciones de aleatoriedad e independencia se asumen de acuerdo con las condiciones de la recolección de los datos. Por otro lado, la asunción de normalidad se puede verificar mediante la prueba de Shapiro Wilk para normalidad con la función 'shapiro.test()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
x	Vector conteniendo las puntuaciones.

Tabla 3.11 - Argumentos de la función 'shapiro.test()' en R

Para concluir, si el p-valor arrojado resulta ser más grande que el nivel de significancia seleccionado, se indica que los tratamientos no tienen efecto sobre la variable respuesta y que la varianza es debido a la misma varianza dentro de los tratamientos o al error experimental. Por otro lado, si el p-valor es menor al nivel de significancia, se rechazaría la hipótesis nula y se concluiría que sí existe diferencia significativa en al menos uno de las muestras y el control.

Cabe destacar que al concluir que existe una diferencia entre las muestras, no se está sugiriendo que, por ejemplo, la muestra B y C obtendrían la misma diferencia si se evaluaran directamente entre ellas, sino que al menos una se considera diferente de la muestra control.

3.4 Prueba de los rangos con signo de Wilcoxon

La prueba de los rangos con signo de Wilcoxon (también conocida como la prueba de Wilcoxon), es la versión no paramétrica equivalente a la prueba T de Student para muestras pareadas comentada anteriormente. Esta prueba verifica la diferencia entre medianas, o más bien, compara las distribuciones de las puntuaciones, pero asume que la dispersión de las distribuciones es la misma, y que, de haber alguna diferencia, sería en el valor de las medianas (O'Mahony, 1986).

La lógica detrás de la prueba de los rangos con signo de Wilcoxon es que bajo la asunción de que dos distribuciones son iguales, se esperaría que la mitad de las diferencias entre las puntuaciones sean negativas, y que la otra mitad sean positivas (Mendenhall, Beaver, & Beaver, 2013); por tanto, para la prueba de diferencia del control (DFC), la hipótesis nula sería que la mediana de las diferencias entre las puntuaciones es igual a 0 (distribuciones idénticas), y la hipótesis alternativa sería que la mediana de las diferencias es diferente de 0 (distribución varía en ubicación).

Para la prueba bilateral con $n < 25$, el estadístico de la prueba se calcula como sigue:

1. Calcular las diferencias $(x_1 - x_2)$ para cada uno de los n pares.
2. Clasificar los valores absolutos de las diferencias, asignando 1 al más pequeño.
3. Calcular T^- , la suma de los rangos para las diferencias negativas y calcular T^+ , la suma de los rangos para las diferencias positivas.

El estadístico sería el menor valor de las cantidades T^- y T^+ calculadas (Wackerly, Mendenhall, & Scheaffer, 2008). Por otro lado, para la prueba bilateral con $n \geq 25$, el estadístico de la prueba se aproxima a la distribución normal y se calcula como sigue:

$$z = \frac{T^+ - [n(n+1)/4]}{\sqrt{[n(n+1)(2n+1)]/24}} \quad (3.4.1)$$

En la práctica, para el cálculo del estadístico y el p-valor de la prueba se utilizará la función 'wilcox.test()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
x	Vector conteniendo las puntuaciones de diferencias para la muestra control.
y	Vector conteniendo las puntuaciones de diferencias para la muestra en evaluación, estos valores deben estar pareados con los introducidos en el argumento 'x'.
alternative	Este argumento indica la hipótesis alternativa y debería indicarse como "two.sided" para calcular el p-valor considerando ambas colas, por lo que esta sería la opción adecuada para la hipótesis alternativa bilateral de la prueba DFC.
paired	Argumento para indicar si se realizará una prueba de wilcoxon para muestras pareadas. Para la prueba DFC debe indicarse como "TRUE".

Tabla 3.12 - Argumentos de la función 'wilcox.test()' en R

En términos de conclusiones, se rechazaría la hipótesis nula si el p-valor arrojado resulta ser más pequeño que el nivel de significancia seleccionado, entonces, se concluiría que sí existe diferencia significativa entre la muestra y el control.

3.5 Prueba de Friedman

La prueba de Friedman para diseño con bloque aleatorio (conocido también como Friedman's ANOVA), es la versión no paramétrica equivalente al análisis de la varianza de un factor con diseño de bloque aleatorio. Esta prueba compara las distribuciones de k tratamientos (productos) dispuestos en b bloques (jueces) (Mendenhall, Beaver, & Beaver, 2013).

La hipótesis nula de la prueba indica que los k tratamientos (productos) siguen la misma distribución, y por tanto, las diferencias de sus medianas es igual a 0. Por otro lado, la hipótesis alternativa indica que al menos uno de los k tratamientos (productos) posee una mediana distinta.

Bajo la hipótesis nula, el estadístico de la prueba se aproxima mediante una distribución chi cuadrado con $(k - 1)$ grados de libertad, y su cálculo es como sigue:

$$F_r = \frac{12}{bk(k+1)} \sum T_i^2 - 3b(k+1) \quad (3.5.1)$$

Siendo T_1, T_2, \dots, T_k la suma de los rangos para cada tratamiento (producto).

En la práctica, para el cálculo del estadístico y el p-valor de la prueba se utilizará la función 'friedman.test()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
y	Vector o matriz de datos utilizado cuando la data es recolectada en formato ancho.
formula	<p>En el caso de que la data esté recolectada en el formato largo, se expresa la fórmula para el análisis de los datos, de la forma:</p> $a \sim k b$ <p>donde a, k y b, son las variables conteniendo los resultados de las puntuaciones, los tratamientos (productos) y los bloques (jueces), respectivamente.</p> <p>En caso de que se encuentre en el formato ancho, se asume que cada fila va a representar un bloque (juez) y cada columna va a representar un tratamiento (producto), por lo que no sería necesario especificar la fórmula.</p>
data	Matriz conteniendo las variables de la formula cuando la data está recolectada en formato largo.

Tabla 3.13 - Argumentos de la función 'friedman.test()' en R

Para concluir, si el p-valor es menor al nivel de significancia, se rechazaría la hipótesis nula y se concluiría que sí existe diferencia significativa en al menos uno de las muestras y el control.

3.6 Prueba de chi cuadrado (χ^2)

Las pruebas presentadas en la Tabla 3.3 comparan los resultados de la proporción de jueces que identificaron la respuesta correcta contra una proporción de valor fijo que representa la probabilidad de acertar al azar de acuerdo con el método utilizado (por ejemplo, 1/3 para la prueba triangular y 1/2 para la prueba dúo-trío). En el caso de las pruebas Igual-Diferente y “A”-“No A” no se identifica una probabilidad de acertar al azar, y el análisis de los resultados se basa en las pruebas chi cuadrado χ^2 de Pearson para homogeneidad e independencia, y la prueba chi cuadrado χ^2 de McNemar para proporción correlacionada. (Bi, Sensory Discrimination Tests and Measurements, 2015)

Tal como su nombre sugiere, bajo la hipótesis nula, el estadístico de las pruebas χ^2 de Pearson y McNemar siguen asintóticamente una distribución chi cuadrado con 1 grado de libertad, $X^2 \sim \chi_1^2$.

La distribución chi cuadrado χ^2 con k grados de libertad es la distribución de una suma de los cuadrados de k variables aleatorias normales estándar independientes. Es utilizada para la prueba de hipótesis sobre la frecuencia de ocurrencia cuando dos variables son independientes y posee las siguientes propiedades (Triola, 2018):

- Los valores de X^2 son no negativos, y la distribución no es simétrica.
- Existe una distribución χ^2 diferente para cada grado de libertad k .

La función de densidad de probabilidad (PDF) es:

$$f_x(x) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{1}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x > 0 \quad (3.6.1)$$

donde Γ es la función gamma.

Siguiendo la misma notación, la función de distribución acumulativa (CDF) viene dada por:

$$F_x(x) = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \quad (3.6.2)$$

donde $\gamma(k, z)$ es la función gamma incompleta.

La selección de la técnica estadística a utilizar para analizar los datos provenientes de una prueba Igual-Diferente y “A”-“No A” depende del tipo de diseño utilizado para la presentación de las muestras a los jueces (Bi, Sensory Discrimination Tests and Measurements, 2015):

Diseño monádico:

Los jueces solo evalúan una muestra (“A” o “No A”) y se determina con antelación la cantidad de jueces que recibirá muestras “A” y muestras “No A”. Los resultados se presentan en una tabla de contingencia 2x2:

		Muestra		Total
		A	No A	
Respuesta	"A"	n_{11}	n_{12}	n_1
	"No A"	n_{21}	n_{22}	$N - n_1$
Total		N_1	N_2	N

Tabla 3.14 - Tabla de contingencia para diseño monádico

Para el análisis de los resultados, se utiliza la prueba chi cuadrado χ^2 de Pearson para homogeneidad.

Diseño mixto:

El número de jueces que van a evaluar muestras "A" y muestras "No A" no se determinan antes de la sesión de evaluación, sino que solo se determina la cantidad de jueces a participar, de ahí cada juez toma una muestra aleatoriamente por lo que se desconoce cuántos jueces tomarán la muestra A o la muestra No A. Los resultados se presentan en una tabla de contingencia 2x2:

		Muestra		Total
		A	No A	
Respuesta	"A"	n_{11}	n_{12}	$n_{11} + n_{12}$
	"No A"	n_{21}	n_{22}	$n_{21} + n_{22}$
Total		$n_{11} + n_{21}$	$n_{12} + n_{22}$	n

Tabla 3.15 - Tabla de contingencia para diseño mixto

Para el análisis de los resultados, se utiliza la prueba chi cuadrado χ^2 de Pearson para independencia.

Diseño pareado:

Todos los jueces evalúan tanto la muestra "A" como la muestra "No A" pero no se les notifica que estarán recibiendo ambas muestras, solo deben identificar si la muestra es "A" o "No A". Los resultados se presentan en una tabla de contingencia 2x2:

		Muestra		Total
		A	No A	
Respuesta	"A"	a	b	
	"No A"	c	d	
Total				N

Tabla 3.16 - Tabla de contingencia para diseño pareado

Para el análisis de los resultados, se utiliza la prueba chi cuadrado χ^2 de McNemar para proporciones correlacionadas

Para utilizar los diseños expuestos anteriormente para la prueba "Igual – Diferente", lo que se expresa como muestra "A" se define como el par de muestras emparejadas (o iguales), y la muestra "No A" se define como el par de muestras no emparejadas (diferentes).

3.6.1 Prueba de chi cuadrado χ^2 de Pearson para homogeneidad

Siguiendo la notación de la Tabla 3.14, el propósito del diseño monádico es verificar si la proporción de respuestas “A” entre los jueces que recibieron muestras “A” es igual a los que recibieron muestras “No A”. Esto sería una comparación estadística entre dos proporciones independientes de dos poblaciones con tamaño muestral N_1 y N_2 mediante la prueba de chi χ^2 de Pearson para homogeneidad (Bi, Sensory Discrimination Tests and Measurements, 2015).

La hipótesis nula sería:

$$H_0: p_A = p_N = \bar{p} \quad (3.6.3)$$

La hipótesis alternativa sería:

$$H_1: p_A > p_N \quad (3.6.4)$$

donde:

p_A es la proporción de respuesta “A” para muestras A.

p_N es la proporción de respuesta “A” para muestras No A.

\bar{p} es el mejor estimado de la probabilidad de respuesta “A”

Para comparar dos proporciones independientes, se puede usar el estadístico de la prueba de chi cuadrado para homogeneidad:

$$X_p^2 = \sum_{j=1}^2 \sum_{i=1}^2 \frac{(n_{ij} - \widehat{E}_{ij})^2}{\widehat{E}_{ij}} \quad (3.6.5)$$

donde:

n_{ij} son las frecuencias observadas en las celdas de la tabla de contingencia 2x2.

E_{ij} son las frecuencias esperadas en las celdas de la tabla de contingencia 2x2.

Dado que el mejor estimado de la probabilidad de respuesta “A” es $\bar{p} = \frac{(n_{11}+n_{12})}{N}$ y el mejor estimado de la probabilidad de respuesta “No A” es $1 - \bar{p} = \frac{(n_{21}+n_{22})}{N}$, entonces, el mejor estimado de las frecuencias esperadas en las celdas son:

		Muestra	
		A	No A
Respuesta	"A"	$E_{11} = \frac{N_1(n_{11} + n_{12})}{N}$	$E_{12} = \frac{N_2(n_{11} + n_{12})}{N}$
	"No A"	$E_{21} = \frac{N_1(n_{21} + n_{22})}{N}$	$E_{22} = \frac{N_2(n_{21} + n_{22})}{N}$

Tabla 3.17 - Frecuencias esperadas para diseño monádico

Se propone la corrección de Yates para compensar el hecho de que se utiliza una distribución continua (χ^2) para aproximar la distribución discreta del estadístico de chi cuadrado, ya que esta no se aproxima correctamente cuando la frecuencia esperada de una o más celdas de la tabla de contingencia 2x2 es muy pequeña, convencionalmente menor a 5. (Yates, 1934). La corrección de continuidad propuesta por Yates consiste en restar 0.5 al valor absoluto de las diferencias entre las frecuencias observadas y las frecuencias esperadas con el propósito de compensar el sesgo hacia arriba del estadístico original, entonces, la ecuación (3.6.5) quedaría como sigue:

$$X_p^2 = \sum_{j=1}^2 \sum_{i=1}^2 \frac{(|n_{ij} - \widehat{E}_{ij}| - 0.5)^2}{\widehat{E}_{ij}} \quad (3.6.6)$$

Una vez se define el nivel de significancia de la prueba (convencionalmente se selecciona un α de 0.1, 0.05 ó 0.01), los valores críticos de la prueba se pueden determinar mediante la siguiente ecuación (Triola, 2018):

$$X^2 = \frac{1}{2} \left[\pm z_{\alpha/2} + \sqrt{2k - 1} \right]^2 \quad (3.6.7)$$

Los cálculos se podrían realizar manualmente o utilizando las tablas de valores críticos para predeterminados grados de libertad y nivel de significancia que proveen los libros de texto de estadística y organizaciones como NIST ("National Institute of Standards and Technology"). En adición, se calcula mediante la función 'qchisq()' del paquete base 'stats' de R (R Core Team, 2022), utilizando los siguientes argumentos:

Argumento	Descripción
p	El vector de probabilidades representa el nivel de significancia α seleccionado.
size	Cantidad de observaciones
df	Grados de libertad = 1
lower.tail	Este argumento está indicado como "TRUE" por defecto, lo que calcularía el valor crítico para la cola inferior de los valores. Al indicar "FALSE" se calcularía el valor crítico para la cola superior, por lo que esta sería la opción adecuada para la hipótesis alternativa unilateral $H_1: p_A > p_N$

Tabla 3.18 - Argumentos de la función 'qchisq()' en R

Para el cálculo del estadístico y p-valor de la prueba, se utilizará la función 'chisq.test()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
x	Matriz conteniendo los valores observados.
correct	Argumento para indicar la utilización de la corrección de continuidad de Yates. Este argumento está indicado como "TRUE" por defecto, por lo que se realizaría la corrección para las tablas de contingencia 2x2 con frecuencias esperadas menor a 5. En caso de que no sea necesario, se colocaría este argumento como "FALSE".

Tabla 3.19 - Argumentos de la función 'chisq.test()' en R para diseño monádico

En términos de conclusiones, se rechazaría la hipótesis nula si el p-valor arrojado resulta ser más pequeño que el nivel de significancia seleccionado, entonces, se concluiría que sí existe diferencia significativa entre los dos productos porque la cantidad de jueces que responden "A" no es homogénea entre la presentación de la muestra "A" y la muestra "No A"

3.6.2 Prueba de chi cuadrado χ^2 de Pearson para independencia

Siguiendo la notación de la Tabla 3.15, el propósito del diseño mixto es verificar si la respuesta "A" o "No A" está asociada a la presentación de la muestra A o No A. Esto sería una prueba estadística sobre la independencia de dos variables, X y Y de una población con tamaño muestral N mediante la prueba de chi χ^2 de Pearson para independencia (Bi, Sensory Discrimination Tests and Measurements, 2015).

Cada variable tiene dos categorías (0,1):

X = 1 indica una respuesta de "A", y X = 0 indica una respuesta de "No A"

Y = 1 indica una muestra A y Y = 0 indica una muestra No A

Entonces, cada uno de los N jueces entran en una de cuatro categorías:

(1,1) indicando que seleccionó la respuesta "A" cuando se le presentó una muestra A.

(1,0) indicando que seleccionó la respuesta "A" cuando se le presentó una muestra No A.

(0,1) indicando que seleccionó la respuesta "No A" cuando se le presentó una muestra A.

(0,0) indicando que seleccionó la respuesta "No A" cuando se le presentó una muestra No A.

La hipótesis nula es que las dos variables X (respuestas) y Y (muestras) son independientes una de la otra, lo que indicaría que la probabilidad de cada celda es igual al producto de las probabilidades de su respectiva fila y columna. La expresión sería como sigue:

$$H_0: p_{ij} = p_i \cdot p_j \quad (3.6.8)$$

La hipótesis alternativa es que existe una relación entre las muestras y las respuestas:

$$H_1: p_{ij} \neq p_i \cdot p_j \quad (3.6.9)$$

Cuanto mayor sea la diferencia entre p_{ij} y $p_i \cdot p_j$, ó entre n_{ij} y \hat{E}_{ij} mayor es la evidencia contra la hipótesis nula.

El estadístico de la prueba de χ^2 de Pearson para independencia es el mismo que el indicado para la prueba de χ^2 de Pearson para homogeneidad en la ecuación (3.6.5) y (3.6.6), por lo que de igual manera se puede utilizar la función 'chisq.test()' del paquete base 'stats' de R (R Core Team, 2022) para el cálculo del estadístico y p-valor de la prueba:

Argumento	Descripción
x	Tabla de contingencia conteniendo los valores observados.
correct	Argumento para indicar la utilización de la corrección de continuidad de Yates. Este argumento está indicado como "TRUE" por defecto, por lo que se realizaría la corrección para las tablas de contingencia 2x2 con frecuencias esperadas menor a 5. En caso de que no sea necesario, se colocaría este argumento como "FALSE".

Tabla 3.20 - Argumentos de la función 'chisq.test()' en R para diseño mixto

En términos de conclusiones, se rechazaría la hipótesis nula si el p-valor arrojado resulta ser más pequeño que el nivel de significancia seleccionado, entonces, se concluiría que sí existe diferencia significativa entre los dos productos porque la cantidad de jueces que responden "A" y los que responden "No A" está relacionado con la presentación de la muestra "A" y la muestra "No A".

A pesar de que se utiliza el mismo estadístico de prueba para el diseño monádico y el diseño mixto, las conclusiones dadas por la prueba de independencia solo son válidas si se cumple la asunción teórica de que solo se fija la cantidad total de jueces antes de realizar la sesión de evaluación.

3.6.3 Prueba de chi cuadrado χ^2 de McNemar para proporción correlacionada

En la prueba de diseño pareado, un mismo juez evalúa una muestra "A" y otra muestra "No A" por lo que la comparación entre las proporciones no es independiente, y se recomienda el uso de la prueba chi cuadrada χ^2 de McNemar para proporción correlacionada (Bi, Sensory Discrimination Tests and Measurements, 2015).

La hipótesis nula es que los jueces tienen la misma probabilidad de indicar una respuesta "A" para una muestra A y para una muestra No A, se expresaría como sigue:

$$H_0: P_A = P_N \quad (3.6.10)$$

Siguiendo la notación de la Tabla 3.16, las probabilidades P_A y P_N se estimarían como sigue:

$$\hat{P}_A = \frac{a+c}{N} \text{ y } \hat{P}_N = \frac{a+b}{N}$$

La hipótesis alternativa se expresaría como sigue:

$$H_0: P_A > P_N \quad (3.6.11)$$

El estadístico para la prueba chi cuadrado χ^2 de McNemar para proporción correlacionada se calcula mediante la siguiente ecuación:

$$X_M^2 = \frac{(b - c)^2}{b + c} \quad (3.6.12)$$

Este estadístico no se aproxima correctamente a la distribución de chi cuadrada χ^2 para valores pequeños, por lo que se realiza una corrección por continuidad propuesta por Edwards:

Para $b + c \geq 30$:

$$X_M^2 = \frac{(b - c)^2}{(b + c + 1)} \quad (3.6.13)$$

Para $8 \leq b + c < 30$:

$$X_M^2 = \frac{(|b - c| - 1)^2}{(b + c + 1)} \quad (3.6.14)$$

En la práctica, para el cálculo del estadístico y el p-valor de la prueba se utilizará la función 'mcnemar.test ()' del paquete base 'stats' de R (R Core Team, 2022), con los siguientes argumentos:

Argumento	Descripción
x	Matriz de contingencia conteniendo los valores observados.
correct	Argumento para indicar la utilización de la corrección de continuidad. Este argumento está indicado como "TRUE" por defecto, por lo que se realizaría la corrección para las tablas de contingencia 2x2 con frecuencias esperadas menor a 5. En caso de que no sea necesario, se colocaría este argumento como "FALSE".

Tabla 3.21 - Argumentos de la función 'mcnemar.test ()' en R

En términos de conclusiones, se rechazaría la hipótesis nula si el p-valor arrojado resulta ser más pequeño que el nivel de significancia seleccionado, entonces, se concluiría que sí existe diferencia significativa porque la muestra presentada tiene un efecto significativo sobre la respuesta "A" o "No A".

4 Análisis de la potencia estadística

La potencia de una prueba de hipótesis ($1 - \beta$) se define como la probabilidad de rechazar una hipótesis nula que realmente es falsa, para las pruebas discriminativas de diferencia, sería la probabilidad de observar una diferencia que realmente existe.

El análisis de la potencia estadística es importante en el contexto de análisis sensorial porque los resultados de una prueba con baja potencia pueden ser ambiguos dado que la probabilidad *a priori* de rechazar la hipótesis nula es baja, y, por tanto, podría nunca detectar una diferencia existente. Convencionalmente, una prueba con una potencia mayor a 80% es considerada estadísticamente poderosa (Dattalo, 2008).

El análisis de la potencia ($1 - \beta$) depende de los siguientes parámetros (Cohen, 1988):

- **Nivel de significancia (α):** representa el riesgo de rechazar la hipótesis nula cuando esta es verdadera. Siendo generalmente un valor pequeño (tales como 0.01, 0.05, 0.10) ya que mientras más pequeño sea, el estándar para el rechazo de la hipótesis nula es más riguroso.
- **Precisión de la muestra y tamaño muestral (n):** la precisión de la muestra es qué tanto se espera que represente el valor real de la población. Dependiendo del modelo estadístico, la precisión puede o no estar relacionada a la unidad de medición, el valor poblacional y la forma de la distribución de la población. Sin embargo, siempre estará relacionado con el tamaño muestral. Por tanto, de manera intuitiva se puede concluir que en general, a mayor tamaño muestral, mayor potencia estadística porque mayor es la precisión de la muestra.
- **Magnitud del efecto (ES ó d de Cohen):** representa el grado en el cual la hipótesis nula es falsa, por lo que, la hipótesis nula siempre establece que el tamaño del efecto es 0. Este valor es definido por el experimentador y será discutido en detalle para cada una de las pruebas revisadas en este trabajo.

Considerando la relación entre estos parámetros, se puede realizar el análisis de uno en función del resto, por lo que, este capítulo se centrará en la revisión del análisis del tamaño muestral para sentar las bases de la planificación de los experimentos sensoriales alcanzando la potencia estadística deseada. Esto se puede realizar siguiendo dos enfoques distintos:

1. Fórmulas estadísticas de acuerdo con los lineamientos de Cohen en su obra "*Statistical Power Analysis for the Behavioral Sciences*" (Cohen, 1988) y el uso de la librería 'pwr' de R (Champely, 2022) para la implementación práctica de las fórmulas utilizadas para generar las tablas de análisis de la potencia.
2. Técnicas de simulación estadística para las pruebas que generalmente no poseen tablas predeterminadas para el análisis de la potencia.

Considerando que las fórmulas propuestas por Cohen son genéricas y no están contextualizadas al análisis sensorial, se recomendará usar como guía los resultados presentados en el acápite de técnicas de simulación estadística.

4.1 Cálculo de tamaño muestral mediante fórmulas estadísticas

En este acápite se presentan los métodos para la generación de las tablas de análisis de la potencia estadística (determinando la potencia o el tamaño muestral) de acuerdo con lo expuesto por (Cohen, 1988) de manera que se puedan realizar de manera digital en R e incluir valores que no se encuentran en estas tablas.

4.1.1 Prueba Binomial

La detección de una diferencia entre proporciones no es una función de la diferencia entre estas proporciones, ya que la detectabilidad de esta diferencia depende de a qué nivel de la escala de la proporción se encuentra esta magnitud. Por tanto, las proporciones P son transformadas mediante la siguiente relación:

$$\phi = 2 * \text{arcoseno } \sqrt{P} \quad (4.1.1)$$

donde la magnitud de la diferencia será detectada igualmente independientemente del nivel de la escala de la proporción, y, por tanto, la magnitud de la diferencia se define como sigue:

$$h'_2 = \phi_1 - \phi_2 \text{ (direccional)} \quad (4.1.2)$$

$$h'_2 = |\phi_1 - \phi_2| \text{ (no direccional)} \quad (4.1.3)$$

donde ϕ_1 es la transformación arcoseno de la proporción P_1 , y ϕ_2 es la transformación arcoseno de c (valor constante con el cual se compara la proporción en la hipótesis nula).

Considerando que para las pruebas descritas en el acápite 3.1 solo se evalúa una proporción, solo una muestra contribuye a la varianza del error de muestreo ($1/n$ con respecto a lo que aportaría la prueba para dos proporciones), por lo que, la magnitud de la diferencia queda como sigue:

$$h = h'_2 \sqrt{2} \quad (4.1.4)$$

Considerando que las pruebas discriminativas evaluadas en este trabajo fin de máster no serán realizadas sobre datos existentes de la empresa, sino que serán realizadas para productos de clientes externos, se considerará la posibilidad de no tener data histórica disponible, por tanto, no se dispondrá de la proporción P requerida en la ecuación (4.1.1), y se propone utilizar una magnitud del efecto igual a 0.28 de acuerdo a lo expuesto por Cohen para diferencias “pequeñas” (Cohen, 1988).

Para el cálculo de la potencia estadística, Cohen propone la siguiente ecuación:

$$z_{1-b} = h \sqrt{\frac{n}{2}} - z_{1-a} \quad (4.1.5)$$

donde z se corresponde a los percentiles de la curva normal para la potencia y nivel de significancia dado.

Al despejar la ecuación (4.1.5) para n , el tamaño muestral se calcula como sigue:

$$n = 2 \left(\frac{z_{1-\alpha} + z_{1-\beta}}{h} \right)^2 \quad (4.1.6)$$

Para fines prácticos, en este trabajo fin de máster se utilizará la función 'pwr.p.test' del paquete 'pwr' (Champely, 2022) de R (R Core Team, 2022). Para el cálculo del tamaño muestral, los argumentos de la función se completarían como sigue:

Argumento	Descripción
h	Magnitud del efecto (d de Cohen), por convención se colocará como 0.28 para identificar diferencias “pequeñas”.
n	En este argumento se indicaría el tamaño muestral a utilizar para el cálculo de la potencia. En este caso, no se especificará o se colocará como “NULL” para indicar que este será el parámetro a calcular.
sig.level	Representa el nivel de significancia α seleccionado.
power	Potencia estadística deseada, convencionalmente no debería ser menor a 0.80.
alternative	Se especifica la hipótesis alternativa de acuerdo con lo expuesto en la Tabla 3.3., indicándose como “two.sided” para las pruebas bilaterales y “greater” para las pruebas unilaterales hacia la derecha.

Tabla 4.1 - Argumentos de la función 'pwr.p.test' en R

Para la potencia estadística se consideró 80% como el mínimo aceptable y se evaluaron incrementos de un 5% hasta alcanzar la potencia de 95%. En la Tabla 4.2 se presenta el tamaño muestral total para alcanzar los valores especificados de magnitud del efecto, potencia estadística y nivel de significancia, considerando una prueba binomial unilateral.

Magnitud del efecto	Potencia estadística	Nivel de significancia		
		α 0.01	α 0.05	α 0.10
0.10	0.80	1004	619	451
	0.85	1131	719	538
	0.90	1302	857	657
	0.95	1578	1083	857
0.20	0.80	251	155	113
	0.85	283	180	135
	0.90	326	215	165
	0.95	395	271	215
0.28	0.80	129	79	58
	0.85	145	92	69
	0.90	167	110	84
	0.95	202	139	110

Tabla 4.2 - Tamaño muestral - Prueba binomial unilateral

El código R utilizado se presenta en el anexo #1.

4.1.2 Prueba T de Student

Dado el nivel de significancia y la potencia estadística requerida, es necesario determinar la magnitud del efecto para calcular el mínimo tamaño muestral requerido. Esto se logra estandarizando el tamaño del efecto expresado en la unidad de medida de la variable dependiente dividiéndola por la desviación estándar común de las medidas en sus respectivas poblaciones:

$$d = \frac{m_A - m_B}{\sigma} \quad (4.1.7)$$

donde m_A y m_B son las medias poblacionales expresada en la unidad original, σ es la desviación estándar de cualquiera de las poblaciones (asumidas iguales) y d es la magnitud del efecto de Cohen.

Considerando que las pruebas discriminativas evaluadas en este trabajo fin de máster, no serán realizadas sobre datos existentes de la empresa, sino que serán realizadas para productos de clientes externos, se considerará la posibilidad de no tener data histórica disponible, por tanto, no se dispondrá del numerador ni denominador de la ecuación (4.1.7), y se propone utilizar una magnitud del efecto igual a 0.20 de acuerdo a lo expuesto por Cohen para diferencias “pequeñas” (Cohen, 1988).

Para el cálculo de la potencia estadística, Cohen la aproximación dada por (Dixon & Massey, 1957):

$$z_{1-b} = \frac{d(n-1)\sqrt{2n}}{2(n-1) + 1.21(z_{1-a} - 1.06)} - z_{1-a} \quad (4.1.8)$$

donde:

z_{1-b} es el percentil de la curva normal unitaria de la potencia.

z_{1-a} es el percentil de la curva normal unitaria del nivel de significancia.

d es la magnitud del efecto.

n es el tamaño muestral.

Por otro lado, el tamaño muestral proviene de las tablas provistas por (Owen, 1965) para el parámetro de no centralidad de la prueba T de Student, δ , como una función de los grados de libertad a y b :

$$\delta = d \sqrt{\frac{n}{2}} \quad (4.1.9)$$

y despejando para el tamaño muestral:

$$n = \frac{2\delta^2}{d^2} \quad (4.1.10)$$

Para fines prácticos, en este trabajo fin de máster se utilizará la función ‘pwr.t.test’ del paquete ‘pwr’ (Champely, 2022) de R (R Core Team, 2022). Para el cálculo del tamaño muestral, los argumentos de la función se completarían como sigue:

Argumento	Descripción
n	En este argumento se indicaría el tamaño muestral a utilizar para el cálculo de la potencia. En este caso, no se especificará o se colocará como "NULL" para indicar que este será el parámetro a calcular.
d	Magnitud del efecto (d de Cohen), por convención se colocará como 0.2 para identificar diferencias "pequeñas".
sig.level	Representa el nivel de significancia α seleccionado.
power	Potencia estadística deseada, convencionalmente no debería ser menor a 0.80.
type	Tipo de prueba T de Student. Se colocará como "paired" para la prueba Diferente del control (DFC) de acuerdo con lo expuesto en el acápite 3.2.
alternative	Se especifica la hipótesis alternativa, indicándose como "two.sided" para considerar ambas colas de acuerdo con lo expuesto en el acápite 3.2.

Tabla 4.3 - Argumentos de la función 'pwr.t.test' en R

En la Tabla 4.4 se presenta el tamaño muestral requerido de cada grupo para alcanzar los valores especificados de magnitud del efecto, potencia estadística y nivel de significancia, considerando una prueba T de Student bilateral.

Magnitud del efecto	Potencia estadística	Nivel de significancia		
		α 0.01	α 0.05	α 0.10
0.10	0.80	2338	1571	1238
	0.85	2612	1797	1439
	0.90	2978	2103	1714
	0.95	3565	2600	2166
0.20	0.80	586	394	310
	0.85	655	450	361
	0.90	746	527	429
	0.95	893	651	542
0.28	0.80	300	202	159
	0.85	335	231	185
	0.90	382	270	220
	0.95	457	333	277

Tabla 4.4 - Tamaño muestral - Prueba T de Student bilateral

El código R utilizado se presenta en el anexo #1.

4.1.3 ANOVA

Para el análisis de la varianza con igual tamaño muestral n , la magnitud del efecto se calcula como sigue:

$$f = \frac{\sigma_m}{\sigma} \quad (4.1.11)$$

donde:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^k (m_i - m)^2}{k}} \quad (4.1.12)$$

Los valores entre paréntesis son las diferencias de las medias poblacionales de la media de medias, k es la cantidad de jueces y σ es la desviación estándar entre las poblaciones.

Para determinar el valor apropiado de la magnitud del efecto f , se requiere el uso de data histórica o previo conocimiento sobre los productos a probar. Por lo que, se propone utilizar la convención de Cohen para efectos “pequeños”, $f=0.10$.

Para el cálculo de la potencia, Cohen propone la raíz cuadrada de Laubscher de la aproximación normal del estadístico F ratio no central:

$$z_{1-b} = \frac{\sqrt{2(u + \lambda) - \frac{u + 2\lambda}{u + \lambda}} - \sqrt{(2v - 1) \frac{u F_c}{v}}}{\sqrt{\frac{u F_c}{v} + \frac{u + 2\lambda}{u + \lambda}}} \quad (4.1.13)$$

Donde el parámetro de no centralidad es:

$$\lambda = f^2 n (u + 1) \quad (4.1.14)$$

F_c es el valor crítico para el estadístico F ratio y u los grados de libertad del numerador.

Los grados de libertad del denominador es:

$$v = (u + 1)(n - 1) \quad (4.1.15)$$

Para el cálculo del tamaño muestral, Cohen propone:

$$n = \left(\frac{\phi}{f} \right)^2 \quad (4.1.16)$$

donde $\phi = \sqrt{\frac{\lambda}{u+1}}$

Para fines prácticos, en este trabajo fin de máster se utilizará la función 'pwr.anova.test' del paquete 'pwr' (Champely, 2022) de R (R Core Team, 2022). Para el cálculo del tamaño muestral, los argumentos de la función se completarían como sigue:

Argumento	Descripción
k	Número de grupos. En este argumento se indicaría el tamaño muestral a utilizar para el cálculo de la potencia. En este caso, no se especificará o se colocará como "NULL" para indicar que este será el parámetro a calcular.
n	Número de observaciones por grupo, se colocaría como 1 ya que un juez no repite muestras.
f	Magnitud del efecto (d de Cohen), por convención se colocará como 0.10 para identificar diferencias "pequeñas".
sig.level	Representa el nivel de significancia α seleccionado.
power	Potencia estadística deseada, convencionalmente no debería ser menor a 0.80.

Tabla 4.5 - Argumentos de la función 'pwr.anova.test' en R

En la Tabla 4.6 se presenta el tamaño muestral de cada grupo para alcanzar los valores especificados de magnitud del efecto, potencia estadística y nivel de significancia, y considerando tres productos o tratamientos ($k = 3$) para el análisis de la varianza (ANOVA).

Magnitud del efecto	Potencia estadística	Nivel de significancia		
		α 0.01	α 0.05	α 0.10
0.10	0.80	465	323	258
	0.85	515	366	297
	0.90	583	423	350
	0.95	690	516	435
0.20	0.80	118	82	66
	0.85	130	93	75
	0.90	147	107	88
	0.95	174	130	110
0.28	0.80	61	42	34
	0.85	68	48	39
	0.90	76	55	46
	0.95	90	67	57

Tabla 4.6 - Tamaño muestral - ANOVA

El código R utilizado se presenta en el anexo #1.

4.1.4 Prueba χ^2 de Pearson

Recordando que la prueba chi cuadrado de Pearson para homogeneidad sigue la misma formulación matemática que la prueba chi cuadrado de Pearson para independencia, en este acápite se unificará el análisis de la potencia estadística ya que en términos de tamaño muestral serían los mismos resultados (suficiente evidencia de la diferencia entre productos) aunque la interpretación de los resultados sería ligeramente distinta.

La magnitud del efecto w viene dada por las frecuencias relativas de las categorías conjuntas. Para cada celda hay dos proporciones una dada por la hipótesis nula y otra dada por la hipótesis alternativa, entonces, la magnitud del efecto mide la discrepancia entre estas proporciones pareadas:

$$w = \sqrt{\sum_{i=1}^m \frac{(p_{1i} - p_{0i})^2}{p_{0i}}} \quad (4.1.17)$$

donde:

p_{0i} es la proporción en la celda i postulada por la hipótesis nula.

p_{1i} es la proporción en la celda i postulada por la hipótesis alternativa y refleja el efecto de dicha celda.

m es la cantidad de celdas.

Para determinar el valor apropiado de la magnitud del efecto w , se requiere el uso de data histórica o previo conocimiento sobre los productos a probar. Por lo que, se propone utilizar la convención de Cohen para efectos “pequeños”, $f=0.10$.

Para el cálculo de la potencia, Cohen propone el uso de las tablas de la distribución acumulativa chi cuadrado no central provistas por Haynam, Govindarajulu y Leone. Al proveer la potencia como una función del parámetro de no centralidad de la distribución chi cuadrada λ (lambda), la relación entre el parámetro de no centralidad λ y la magnitud del efecto w es como sigue:

$$\lambda = w^2 N \quad (4.1.18)$$

Despejando para N , el tamaño muestral se calcula como sigue:

$$N = \frac{\lambda}{w^2} \quad (4.1.19)$$

Para fines prácticos, en este trabajo fin de máster se utilizará la función ‘pwr.chisq.test’ del paquete ‘pwr’ (Champely, 2022) de R (R Core Team, 2022). Para el cálculo del tamaño muestral, los argumentos de la función se completarían como sigue:

Argumento	Descripción
w	Magnitud del efecto (d de Cohen), por convención se colocará como 0.10 para identificar diferencias “pequeñas”.
sig.level	Representa el nivel de significancia α seleccionado.
power	Potencia estadística deseada, convencionalmente no debería ser menor a 0.80.

Tabla 4.7 - Argumentos de la función 'pwr.chisq.test' en R

En la Tabla 4.8 se presenta el tamaño muestral de cada grupo para alcanzar los valores especificados de magnitud del efecto, potencia estadística y nivel de significancia, y considerando un grado de libertad para la prueba chi cuadrado de pearson para homogeneidad e independencia.

Magnitud del efecto	Potencia estadística	Nivel de significancia		
		α 0.01	α 0.05	α 0.10
0.10	0.80	1168	785	619
	0.85	1305	898	719
	0.90	1488	1051	857
	0.95	1782	1300	1083
0.20	0.80	292	197	155
	0.85	327	225	180
	0.90	372	263	215
	0.95	446	325	271
0.28	0.80	149	101	79
	0.85	167	115	92
	0.90	190	135	110
	0.95	228	166	139

Tabla 4.8 - Tamaño muestral - ANOVA

El código R utilizado se presenta en el anexo #1.

4.2 Cálculo de tamaño muestral mediante técnicas de simulación

Para el cálculo del tamaño muestral de las técnicas estadísticas no cubiertas por Cohen, se propone realizar de 5,000 a 10,000 simulaciones del escenario que se correspondería a una hipótesis alternativa de cada prueba para obtener la proporción de veces que se rechaza la hipótesis nula correctamente, y, por ende, determinar la potencia estadística asociada a dicho nivel de significancia, tamaño muestral y magnitud del efecto previamente establecido.

Considerando que no se posee data histórica para realizar una hipótesis sobre la magnitud del efecto, para todos los escenarios expuestos en este acápite se considerará una magnitud de 0.20, convención arbitraria pero generalmente aceptada para identificar diferencias “pequeñas” (Bausell & Li, 2002). Los tamaños muestrales a evaluar se realizarán en incrementos en el tamaño muestral hasta alcanzar una potencia aproximada a 1.

Cabe destacar que este análisis se ha realizado para fines representativos ya que se está haciendo una asunción sobre la naturaleza de la diferencia entre las puntuaciones recibidas por los jueces, por lo que, los resultados mostrados en las tablas de este acápite solo se deben seguir como guía sobre qué potencia esperar de acuerdo con el tamaño muestral, suponiendo una distribución similar para las puntuaciones de cada caso.

4.2.1 Prueba Binomial

Para simular la hipótesis alternativa, se tomó generaron éxitos (aplicable para representar aciertos en cualquier prueba como triangular, tetrad, etc) mediante una distribución binomial con probabilidad de éxito igual a 0.4. El código R utilizado se presenta en el anexo #1 y los resultados se muestran en la Tabla 4.9:

Tamaño muestral	Potencia estadística		
	α 0.01	α 0.05	α 0.10
25	0.3503	0.5115	0.6625
35	0.5110	0.6389	0.8437
45	0.6153	0.8048	0.8048
55	0.7014	0.8603	0.9131
65	0.7625	0.8893	0.9302
75	0.8168	0.9389	0.9640
85	0.9021	0.9562	0.9819
95	0.9254	0.9806	0.9944
105	0.9447	0.9867	0.9966
115	0.9698	0.9926	0.9969
125	0.9773	0.9938	0.9986
135	0.9870	0.9977	0.9995
145	0.9904	0.9984	0.9995

Tabla 4.9 - Tamaño muestral - Prueba binomial

Como es de esperarse, en la Figura 4.4 se observa que la potencia de 0.8 se alcanza más rápido a medida que aumenta el valor del nivel de significancia. Considerando un nivel de significancia de 0.05, la potencia de 0.8 se alcanza con un tamaño muestral aproximado de 45 jueces, mientras que para el nivel de significancia de 0.01 se requerirían alrededor de 75 jueces.

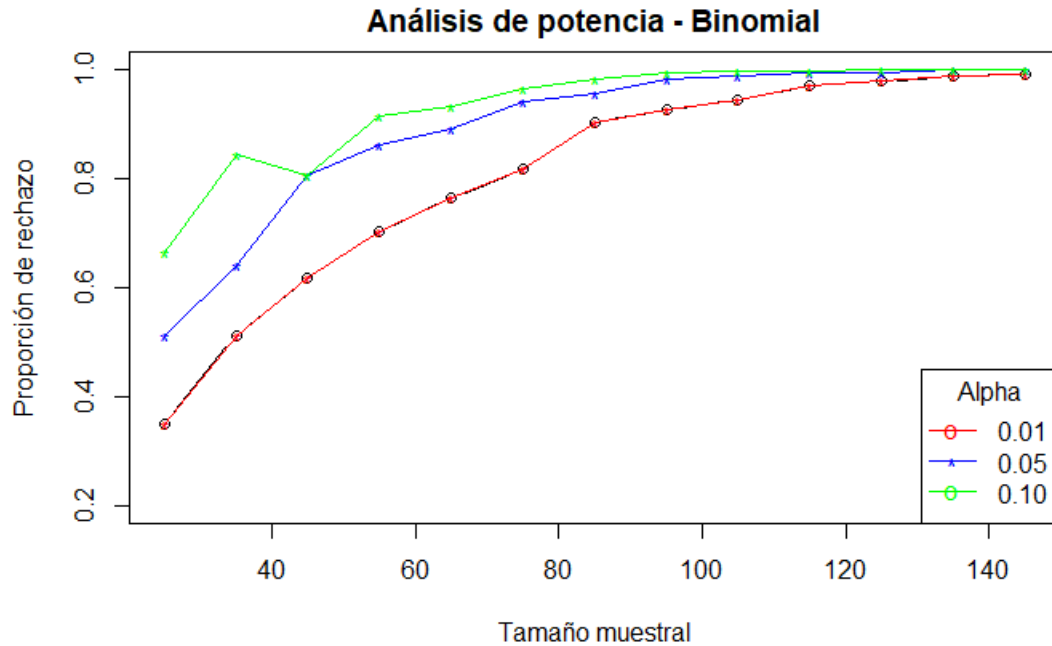


Figura 4.1 - Análisis de potencia - Prueba binomial

4.2.2 Prueba T de Student

Para simular la hipótesis alternativa, se tomó un muestreo aleatorio simple de la escala 0-2 para generar la muestra control y un muestreo aleatorio simple de la escala 0-3 para simular la muestra bajo evaluación con una diferencia “pequeña” contra las puntuaciones de la muestra control. El código R utilizado se presenta en el anexo #1 y los resultados se muestran en la Tabla 4.10:

Tamaño muestral	Potencia estadística		
	α 0.01	α 0.05	α 0.10
n=25	0.1774	0.4022	0.5323
n=35	0.2854	0.5372	0.6699
n=45	0.3896	0.6521	0.7655
n=55	0.5112	0.7576	0.8483
n=65	0.5909	0.8075	0.8866
n=75	0.6970	0.8776	0.9325
n=85	0.7525	0.9050	0.9469
n=95	0.8099	0.9348	0.9689
n=105	0.8586	0.9589	0.9806
n=115	0.8914	0.9687	0.9872
n=125	0.9239	0.9790	0.9917
n=135	0.9398	0.9849	0.9924
n=145	0.9578	0.9903	0.9961

Tabla 4.10 - Tamaño muestral - Prueba de T de Student

Como es de esperarse, en la Figura 4.2 se observa que la potencia de 0.8 se alcanza más rápido a medida que aumenta el valor del nivel de significancia. Considerando un nivel de significancia de 0.05, la potencia de 0.8 se alcanza con un tamaño muestral aproximado de 65 jueces, mientras que para el nivel de significancia de 0.01 se requerirían alrededor de 95 jueces.

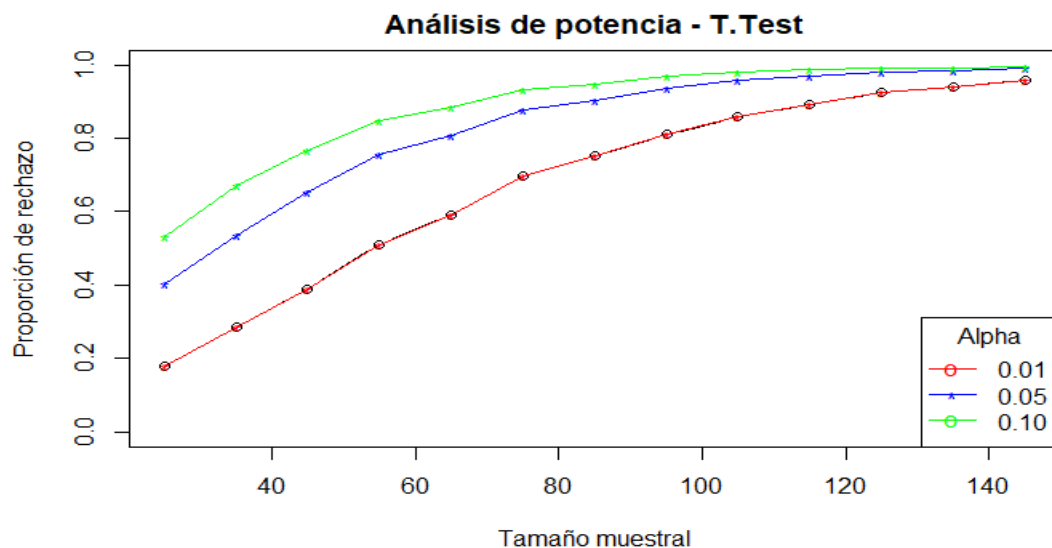


Figura 4.2 - Análisis de potencia - Prueba Wilcoxon

4.2.3 Prueba ANOVA

Para simular la hipótesis alternativa, se tomó un muestreo aleatorio simple de la escala 0-2 para generar la muestra control, un muestreo aleatorio simple de la escala 0-2 para simular la primera muestra bajo evaluación y una escala 0-3 para la segunda muestra en evaluación. Esto para representar una diferencia “pequeña” contra las puntuaciones de la muestra control. El código R utilizado se presenta en el anexo #1 y los resultados se muestran en la Tabla 4.10:

Tamaño muestral	Potencia estadística		
	α 0.01	α 0.05	α 0.10
n=25	0.2370	0.4558	0.5703
n=35	0.3653	0.6005	0.7095
n=45	0.5016	0.7172	0.8073
n=55	0.6068	0.8100	0.8788
n=65	0.7062	0.8675	0.9200
n=75	0.7834	0.9089	0.9477
n=85	0.8410	0.9426	0.9696
n=95	0.8849	0.9621	0.9813
n=105	0.9179	0.9757	0.9878
n=115	0.9449	0.9837	0.9917
n=125	0.9642	0.9912	0.9966
n=135	0.9795	0.9941	0.9980
n=145	0.9841	0.9948	0.9980

Tabla 4.11 - Tamaño muestral - Prueba de T ANOVA

Como es de esperarse, en la Figura 4.3 se observa que la potencia de 0.8 se alcanza más rápido a medida que aumenta el valor del nivel de significancia. Considerando un nivel de significancia de 0.05, la potencia de 0.8 se alcanza con un tamaño muestral aproximado de 55 jueces, mientras que para el nivel de significancia de 0.01 se requerirían alrededor de 85 jueces.

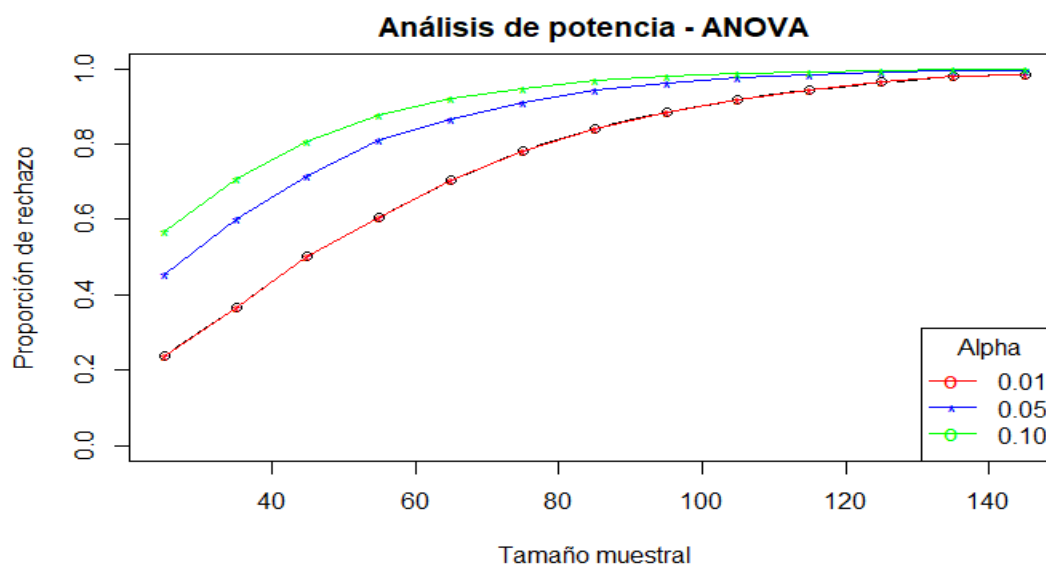


Figura 4.3 - Análisis de potencia - Prueba ANOVA

4.2.4 Prueba de Wilcoxon

Para simular la hipótesis alternativa, se tomó un muestreo aleatorio simple de la escala 0-2 para generar la muestra control y un muestreo aleatorio simple de la escala 0-3 para simular la muestra bajo evaluación con una diferencia “pequeña” contra las puntuaciones de la muestra control. El código R utilizado se presenta en el anexo #1 y los resultados se muestran en la Tabla 4.12:

Tamaño muestral	Potencia estadística		
	α 0.01	α 0.05	α 0.10
n=25	0.1354	0.3725	0.5151
n=35	0.2425	0.5094	0.6472
n=45	0.3452	0.6271	0.7446
n=55	0.4710	0.7354	0.8298
n=65	0.5490	0.7858	0.8693
n=75	0.6551	0.8569	0.9206
n=85	0.7133	0.8873	0.9376
n=95	0.7775	0.9202	0.9599
n=105	0.8268	0.9475	0.9754
n=115	0.8674	0.9591	0.9815
n=125	0.9050	0.9724	0.9864
n=135	0.9243	0.9796	0.9892
n=145	0.9447	0.9857	0.9943

Tabla 4.12 - Tamaño muestral - Prueba de Wilcoxon

Como es de esperarse, en la Figura 4.4 se observa que la potencia de 0.8 se alcanza más rápido a medida que aumenta el valor del nivel de significancia. Considerando un nivel de significancia de 0.05, la potencia de 0.8 se alcanza con un tamaño muestral aproximado de 75 jueces, mientras que para el nivel de significancia de 0.01 se requerirían alrededor de 105 jueces.

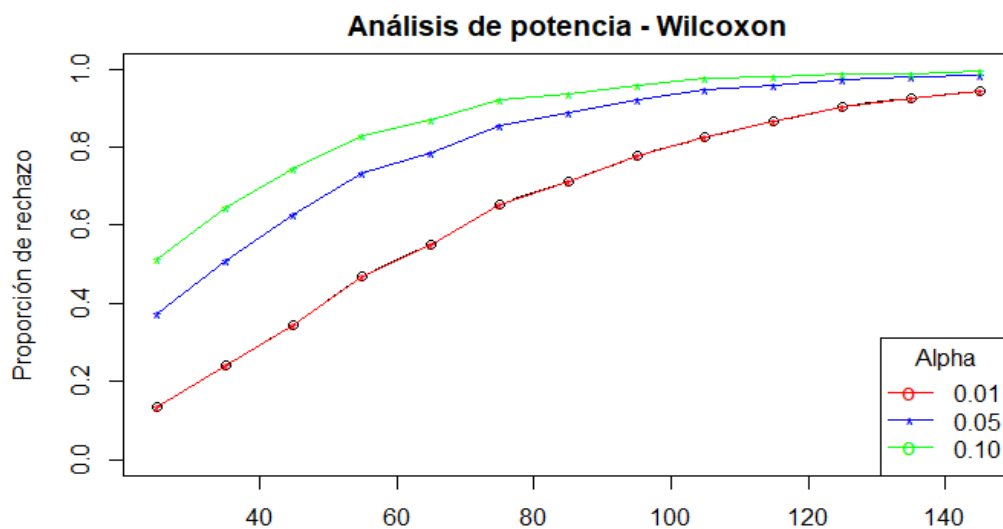


Figura 4.4 - Análisis de potencia - Prueba Wilcoxon

4.2.5 Prueba de Friedman

Para simular la hipótesis alternativa en la prueba Friedman, se tomó un muestreo aleatorio simple de la escala 0-2 para generar la muestra control y un muestreo aleatorio simple de la escala 0-3 para simular la muestra bajo evaluación con una diferencia “pequeña” contra las puntuaciones de la muestra control. La data se generó en formato ancho, por lo que cada fila representa un juez y cada columna un producto (incluyendo la muestra control) de acuerdo con lo expuesto en la Tabla 4.13. El código R utilizado se presenta en el anexo #1 y los resultados obtenidos se muestran debajo:

Tamaño muestral	Potencia estadística		
	α 0.01	α 0.05	α 0.10
n=25	0.3246	0.5936	0.7232
n=35	0.5116	0.7558	0.8530
n=45	0.6772	0.8724	0.9282
n=55	0.7946	0.9286	0.9632
n=65	0.8774	0.9626	0.9846
n=75	0.9266	0.9812	0.9934
n=85	0.9594	0.9900	0.9954
n=95	0.9778	0.9952	0.9982
n=105	0.9886	0.9982	0.9998
n=115	0.9962	0.9996	1.0000

Tabla 4.13 - Tamaño muestral - Prueba de Friedman

Como es de esperarse, en la Figura 4.5 se observa que la potencia de 0.8 se alcanza más rápido a medida que aumenta el valor del nivel de significancia. Considerando un nivel de significancia de 0.05, la potencia de 0.8 se alcanza con un tamaño muestral aproximado de 45, mientras que para el nivel de significancia de 0.01 se requerirían alrededor de 65 jueces.

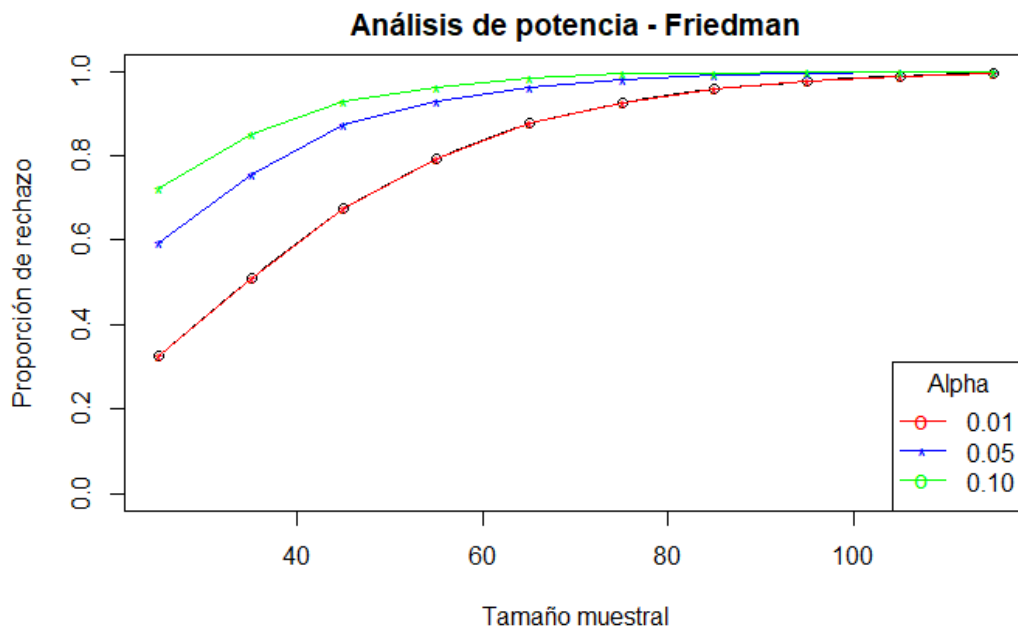


Figura 4.5 - Análisis de potencia - Prueba Friedman

4.3 Conclusiones

Las tablas Tabla 4.2, Tabla 4.4, Tabla 4.6 y Tabla 4.8 se presentan para fines ilustrativos pero se recomienda utilizar las funciones descritas anteriormente para determinar el tamaño muestral requerido durante el proceso de planificación del experimento, o para calcular la potencia estadística a posteriori suponiendo que se tienen limitaciones sobre la cantidad de productos o jueces disponibles.

En caso de poder realizar una prueba piloto o tener históricos sobre los productos a evaluar, se recomienda omitir estas fórmulas ya que son aproximaciones genéricas y no están contextualizadas al análisis sensorial. La recomendación sería utilizar las técnicas de simulación propuestas en el acápite 4.2 para simular la hipótesis alternativa basada en estos datos y calcular la potencia estadística de cada prueba.

Como se puede mostrar en la tabla debajo, para un nivel de significancia de 0.05, el tamaño muestral requerido al calcular la potencia estadística de 0.8 mediante simulación estadística resulta ser mucho más bajo que el calculado con las fórmulas estadísticas clásicas:

Prueba estadística	Potencia estadística ~ 0.8	
	Fórmulas aproximadas	Simulación estadística
Prueba binomial	79 jueces	45 jueces
Prueba T de Student	394 jueces	65 jueces
ANOVA	465 jueces	55 jueces

Los valores de la simulación estadísticas son sujetos a las distribuciones utilizadas para generar los datos, aunque pueden ser más acercadas al contexto del análisis sensorial que las fórmulas genéricas de Cohen, se recomienda utilizar data histórica para identificar la distribución que pueda representar las diferencias de manera más apropiada.

5 Diseño experimental

El diseño experimental consiste en la planificación para el desarrollo de la investigación, incluyendo la recolección de los datos de manera que el análisis realizado posteriormente pueda arrojar resultados y conclusiones válidas. En los capítulos anteriores ya se han abordado aspectos esenciales del diseño experimental en el contexto del análisis sensorial:

1. Definición de la prueba discriminativa a utilizar mediante el diagrama de la Figura 2.11,
2. Cantidad de muestras presentadas a cada juez e hipótesis de la prueba, capítulo 3.
3. Cantidad de evaluaciones requeridas para alcanzar la potencia estadística, capítulo 4.

Para completar el diseño experimental, aún queda por definir en qué orden se presentarían las muestras a cada juez y cuál sería el lineamiento para dicha presentación. Se propone utilizar un diseño experimental completamente aleatorio, donde todas las muestras son asignadas aleatoriamente para minimizar el efecto en la respuesta. Debajo se muestran los posibles órdenes de presentación de las muestras:

Prueba discriminativa		Posible presentación				
Prueba triangular	BAA	ABA	BBA	AAB	BAB	ABB
Prueba tetrad	BBAA	BABA	ABBA	BAAB	ABAB	AABB
Prueba dos de cinco	BBAAA	BABAA	ABBAA	BBBAA	BAABA	ABABA
	BBABA	AABBA	BABBA	ABBB A	BAAAB	ABAAB
	BBAAB	AABAB	BABAB	ABBAB	AAABB	BAABB
	ABABB	AABBB				
Prueba 2-AFC	BA	AB				
Prueba 3-AFC	BAA	ABA	BBA	AAB	BAB	ABB
Prueba dúo-trío	RefA, B, A	RefA, A, A				
Prueba diferente del control (DFC)	ControlA, C, B, A	ControlA, B, C, A	ControlA, C, A, B	ControlA, A, C, B	ControlA, B, A, C	ControlA, A, B, C
Prueba "A" – "No A"- Diseño monádico y mixto	RefA, A	RefA, B				
Prueba "A" – "No A"- Diseño pareado	RefA, A, B	RefA, B, A				
Prueba "Igual –Diferente"- Diseño monádico y mixto	RefA, Igual	RefA, Diferente				
Prueba "Igual –Diferente"- Diseño pareado	RefA, Igual, Diferente	RefA, Diferente, Igual				

Tabla 5.1 - Posibles presentaciones de las muestras

En el anexo #2 se presenta el código R utilizado para determinar las posibles presentaciones para cada prueba discriminativa. Luego, para asignar cada muestra a cada juez de manera aleatoria, se definió la función “generar_diseño”, cuyo código R también está disponible en el anexo #2.

Para utilizar el desarrollo de la función de generar el diseño experimental, se integró dentro de la misma función la generación del dataframe correspondiente a las posibles órdenes de las muestras (definido en el anexo #2). Debajo se muestran los argumentos requeridos para la ejecución de la función “generar_diseño”.

Argumento	Descripción	
prueba	Este argumento indica la prueba discriminativa a realizar. A la fecha, contiene todas las pruebas definidas en el capítulo 2 de este trabajo fin de máster:	
	Valor del argumento	Prueba discriminativa
	"triangular"	Prueba triangular
	"tetrad_esp"	Prueba tetrad especificada
	"tetrad_no"	Prueba tetrad no especificada
	"2of5"	Prueba dos de cinco
	"2AFC"	Prueba elección forzada de 2 alternativas
	"3AFC"	Prueba elección forzada de 3 alternativas
	"duo_trio"	Prueba dúo trio
	"DFC"	Prueba diferente del control
	"AnoA_monadico"	Prueba "A"- "No A" con diseño monádico
	"AnoA_pareado"	Prueba "A"- "No A" con diseño pareado
	"ID_monadico"	Prueba "Igual"- "Diferente" con diseño monádico
"ID_pareado"	Prueba "Igual"- "Diferente" con diseño pareado	
num_jueces	Cantidad de jueces a realizar la evaluación	

Tabla 5.2 - Argumentos de la función ‘generar_diseño ()’ en R

Esta función retornará un dataframe conteniendo el orden en que se presentará el producto a cada juez (para conocimiento del experimentador) y una identificación numérica de la muestra (para presentación al juez). Por ejemplo, una prueba triangular con 5 jueces luciría como sigue:

Juez	Muestra1	Muestra2	Muestra3	id_Muestra1	id_Muestra2	id_Muestra3
<int>	<fctr>	<fctr>	<fctr>	<int>	<int>	<int>
1	A	B	B	632	147	769
2	A	A	B	158	242	774
3	B	B	A	225	257	557
4	A	A	B	135	740	577
5	A	B	B	294	526	961

Figura 5.1 - Representación de diseño experimental para prueba triangular

Considerando que en las pruebas con diseño mixto no se sabe con anticipación cuál producto evaluará cada juez, no se requiere su inclusión en la función para generar el diseño. En adición, si se introduce el valor de una prueba no implementada en la función, se retornará un error indicando que la prueba no fue reconocida.

6 Implementación de las pruebas

Con el fin de asegurar el uso e interpretación correcta de las funciones descritas durante el desarrollo del capítulo 3, en este capítulo se presentan las funciones generadas para la ejecución de cada una de las pruebas.

6.1 Función “prueba_binomial()”.

Se desarrolló la función “prueba_binomial()” para asegurar que la probabilidad de adivinar al azar se esté eligiendo correctamente de acuerdo a los lineamientos de cada una de las pruebas discriminativas que utilizan la prueba binomial. El código R con los comentarios para la ejecución de la prueba binomial se encuentra en el anexo #3.

Debajo se muestran los argumentos requeridos para la ejecución de la función “prueba_binomial()”:

Argumento	Descripción																								
prueba	Este argumento indica la prueba discriminativa a realizar. A la fecha, contiene todas las pruebas definidas en el capítulo 2 de este trabajo fin de máster:																								
	<table><tr><th>Valor del argumento</th><th>Prueba discriminativa</th></tr><tr><td>"triangular"</td><td>Prueba triangular</td></tr><tr><td>"tetrad"</td><td>Prueba tetrad</td></tr><tr><td>"2of5"</td><td>Prueba dos de cinco</td></tr><tr><td>"2AFC"</td><td>Prueba elección forzada de 2 alternativas</td></tr><tr><td>"3AFC"</td><td>Prueba elección forzada de 3 alternativas</td></tr><tr><td>"duo_trio"</td><td>Prueba dúo trio</td></tr><tr><td>"DFC"</td><td>Prueba diferente del control</td></tr><tr><td>"AnoA_monadico"</td><td>Prueba "A"-"No A" con diseño monádico</td></tr><tr><td>"AnoA_pareado"</td><td>Prueba "A"-"No A" con diseño pareado</td></tr><tr><td>"ID_monadico"</td><td>Prueba "Igual"-"Diferente" con diseño monádico</td></tr><tr><td>"ID_pareado"</td><td>Prueba "Igual"-"Diferente" con diseño pareado</td></tr></table>	Valor del argumento	Prueba discriminativa	"triangular"	Prueba triangular	"tetrad"	Prueba tetrad	"2of5"	Prueba dos de cinco	"2AFC"	Prueba elección forzada de 2 alternativas	"3AFC"	Prueba elección forzada de 3 alternativas	"duo_trio"	Prueba dúo trio	"DFC"	Prueba diferente del control	"AnoA_monadico"	Prueba "A"-"No A" con diseño monádico	"AnoA_pareado"	Prueba "A"-"No A" con diseño pareado	"ID_monadico"	Prueba "Igual"-"Diferente" con diseño monádico	"ID_pareado"	Prueba "Igual"-"Diferente" con diseño pareado
	Valor del argumento	Prueba discriminativa																							
	"triangular"	Prueba triangular																							
	"tetrad"	Prueba tetrad																							
	"2of5"	Prueba dos de cinco																							
	"2AFC"	Prueba elección forzada de 2 alternativas																							
	"3AFC"	Prueba elección forzada de 3 alternativas																							
	"duo_trio"	Prueba dúo trio																							
	"DFC"	Prueba diferente del control																							
	"AnoA_monadico"	Prueba "A"-"No A" con diseño monádico																							
	"AnoA_pareado"	Prueba "A"-"No A" con diseño pareado																							
	"ID_monadico"	Prueba "Igual"-"Diferente" con diseño monádico																							
	"ID_pareado"	Prueba "Igual"-"Diferente" con diseño pareado																							
Este argumento será el que realice el enlace con las probabilidades de adivinar al azar asignadas a cada prueba en la Tabla 3.2.																									
exitos	Cantidad de éxitos obtenidos por los jueces, siendo esto cuando el juez acierta en la respuesta correcta (identifica la diferencia asumida por el experimentador).																								
n	Cantidad de ensayos																								
alpha	Nivel de significancia seleccionado. Por default, igual a 0.05.																								

Tabla 6.1 - Argumentos de la función ‘prueba_binomial ()’ en R

6.2 Función “prueba_DFC()”.

Tomando en consideración que se utilizan varias técnicas para la prueba diferente del control, para su aplicación se desarrolló una función que determine cuál debería ser la técnica estadística a utilizar de acuerdo a las propiedades expuestas en el capítulo 3.

El código R con los comentarios para la ejecución de la prueba DFC se encuentra en el anexo #3. Debajo se muestran los argumentos requeridos para la ejecución de la función “prueba_DFC()”:

Argumento	Descripción
Resultados_control	Dataframe incluyendo las puntuaciones para el control oculto.
Resultados_prueba	Dataframe incluyendo las puntuaciones para las muestras en evaluación
alpha	Nivel de significancia seleccionado. Por default, igual a 0.05.

Tabla 6.2 - Argumentos de la función ‘prueba_binomial ()’ en R

6.3 Función “prueba_chisqA()”.

Se desarrolló la función “prueba_chisqA()” para asegurar que se ejecute la prueba chi cuadrado correspondiente de acuerdo con el diseño utilizado para la prueba “A”-“No A” y que se realice la corrección correspondiente cuando las frecuencias esperadas sean menores a 5. El código R con los comentarios para la ejecución de la prueba chi cuadrado se encuentra en el anexo #3.

Debajo se muestran los argumentos requeridos para la ejecución de la función “prueba_chisqA()”

Argumento	Descripción								
datos	Matriz de datos presentados en una tabla de contingencia 2x2.								
diseño	<div><div>Este argumento indica el diseño utilizado para la recolección de los datos de la prueba “A”-“No A”.</div><table><tr><th>Valor del argumento</th><th>Prueba discriminativa</th></tr><tr><td>"monadico"</td><td>Diseño monádico: cada juez evalúa una muestra “A” ó una muestra “No A”. Se realiza la prueba chi cuadrado de Pearson para homogeneidad.</td></tr><tr><td>"mixto"</td><td>Diseño mixto: no se sabe con anticipación la cantidad de muestras “A” y muestras “No A” que evaluará cada juez. Se realiza la prueba chi cuadrado de Pearson para independencia.</td></tr><tr><td>"pareado"</td><td>Diseño pareado: cada juez evalúa una muestra “A” y una muestra “No A”. Se realiza la prueba de chi cuadrado de McNemar para proporción correlacionadas.</td></tr></table></div>	Valor del argumento	Prueba discriminativa	"monadico"	Diseño monádico: cada juez evalúa una muestra “A” ó una muestra “No A”. Se realiza la prueba chi cuadrado de Pearson para homogeneidad.	"mixto"	Diseño mixto: no se sabe con anticipación la cantidad de muestras “A” y muestras “No A” que evaluará cada juez. Se realiza la prueba chi cuadrado de Pearson para independencia.	"pareado"	Diseño pareado: cada juez evalúa una muestra “A” y una muestra “No A”. Se realiza la prueba de chi cuadrado de McNemar para proporción correlacionadas.
Valor del argumento	Prueba discriminativa								
"monadico"	Diseño monádico: cada juez evalúa una muestra “A” ó una muestra “No A”. Se realiza la prueba chi cuadrado de Pearson para homogeneidad.								
"mixto"	Diseño mixto: no se sabe con anticipación la cantidad de muestras “A” y muestras “No A” que evaluará cada juez. Se realiza la prueba chi cuadrado de Pearson para independencia.								
"pareado"	Diseño pareado: cada juez evalúa una muestra “A” y una muestra “No A”. Se realiza la prueba de chi cuadrado de McNemar para proporción correlacionadas.								
alpha	Nivel de significancia seleccionado. Por default, igual a 0.05.								

Tabla 6.3 - Argumentos de la función ‘prueba_chisqA ()’ en R

La función arrojará una conclusión distinta para cada tipo de diseño:

Diseño	Pvalor < nivel de significancia	Pvalor > nivel de significancia
Monádico	Se rechaza la hipótesis nula y se concluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que responden 'A' no es homogénea entre la presentación de las muestras 'A' y 'No A'	no existe evidencia para determinar diferencia significativa entre los productos evaluados
Mixto	Se rechaza la hipótesis nula y se concluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que responden 'A' está relacionado con la presentación de las muestras 'A' y 'No A'	no existe evidencia para determinar diferencia significativa entre los productos evaluados
Pareado	Se rechaza la hipótesis nula y se concluye que existe una diferencia significativa entre los productos ya que la muestra presentada tiene un efecto significativo sobre la respuesta 'A' o 'No A'.	no existe evidencia para determinar diferencia significativa entre los productos evaluados.

Tabla 6.4 - Interpretación de resultados prueba "A"- "No A"

6.4 Función “prueba_chisqID()”.

La función “prueba_chisqID()” está desarrollada de manera idéntica a la función “prueba_chisqA” con la distinción de que las conclusiones vienen dadas para interpretar la presentación de las muestras “Igual”- “Diferente”. De todos modos, en el anexo #3 se provee el código R y debajo los argumentos de la función:

Argumento	Descripción	
datos	Matriz de datos presentados en una tabla de contingencia 2x2.	
diseño	Este argumento indica el diseño utilizado para la recolección de los datos de la prueba “Igual”- “Diferente”.	
	Valor del argumento	Prueba discriminativa
	"monadico"	Diseño monádico: cada juez evalúa una muestra “Igual” ó una muestra “Diferente”. Se realiza la prueba chi cuadrado de Pearson para homogeneidad.
	"mixto"	Diseño mixto: no se sabe con anticipación la cantidad de muestras “Igual” y muestras “Diferente” que evaluará cada juez. Se realiza la prueba chi cuadrado de Pearson para independencia.
	"pareado"	Diseño pareado: cada juez evalúa una muestra “Igual” y una muestra “Diferente”. Se realiza la prueba de chi cuadrado de McNemar para proporción correlacionadas.
alpha	Nivel de significancia seleccionado. Por default, igual a 0.05.	

Tabla 6.5 - Argumentos de la función ‘prueba_chisqID ()’ en R

La función arrojará una conclusión distinta para cada tipo de diseño:

Diseño	Pvalor < nivel de significancia	Pvalor > nivel de significancia
Monádico	Se rechaza la hipótesis nula y se concluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que responden 'Igual' no es homogénea entre la presentación de las muestras 'Igual' y 'Diferente'	No existe evidencia para determinar diferencia significativa entre los productos evaluados
Mixto	Se rechaza la hipótesis nula y se concluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que responden 'Igual' está relacionado con la presentación de las muestras 'Igual' y 'Diferente'	No existe evidencia para determinar diferencia significativa entre los productos evaluados
Pareado	Se rechaza la hipótesis nula y se concluye que existe una diferencia significativa entre los productos ya que la muestra presentada tiene un efecto significativo sobre la respuesta 'Igual' o 'Diferente'.	No existe evidencia para determinar diferencia significativa entre los productos evaluados.

Tabla 6.6 - Interpretación de resultados prueba "Igual"-"Diferente"

7 Simulación de datos y ejecución de las pruebas

Considerando que la empresa TasteLab S. L. presta el servicio de planificación y análisis para la prueba triangular, no se tienen datos para representar todas las pruebas estadísticas evaluadas en este trabajo fin de máster. En adición, las pruebas de análisis sensorial son de especial interés durante el proceso de investigación y desarrollo de los productos, siendo estos procesos confidenciales. Dicho esto, se presentará la simulación de los datos correspondiente a las pruebas adicionales que se pretenden implementar en este trabajo fin de máster con el propósito de ejemplificar el uso de las pruebas desarrolladas en el capítulo anterior.

Para fines de reproducibilidad, se utilizará la semilla “12424” durante la simulación de los datos. Esta simulación constará de los siguientes pasos para cada una de las pruebas:

1. Generación del diseño experimental
2. Identificación de la respuesta correcta de acuerdo con el diseño experimental.
3. Generación de los datos para 60 jueces.
4. Lectura de los datos (comparación de la respuesta simulada vs. la respuesta correcta).
5. Análisis de los datos mediante la prueba correspondiente (interpretación del p-valor).

7.1 Prueba binomial

Para ejemplificar el uso de la función “prueba_binomial()”, se simularon datos de una prueba triangular, definiendo una mayor proporción de respuestas correctas al seleccionar respuestas de la columna “respuestas_correctas” generadas en el punto 2 detallado anteriormente. Referirse al código expuesto en el anexo #4.

El código R utilizado se presenta en el anexo #4. Debajo se muestran los primeros 6 valores (“head”) del dataframe que contiene el diseño experimental y las respuestas simuladas:

Description: df [6 x 10]										
	Juez	Muestra1	Muestra2	Muestra3	id_Muestra1	id_Muestra2	id_Muestra3	respuesta_correcta	respuesta_juez	identificacion_correcta
	<int>	<fctr>	<fctr>	<fctr>	<int>	<int>	<int>	<int>	<int>	<chr>
1	1	A	B	B	353	613	137	353	613	Incorrecta
2	2	A	A	B	958	792	988	988	958	Incorrecta
3	3	B	B	A	476	965	760	760	760	Correcta
4	4	A	A	B	329	455	319	319	319	Correcta
5	5	A	B	B	148	588	424	148	588	Incorrecta
6	6	A	B	B	701	154	396	701	701	Correcta

Figura 7.1 - Resultados simulación de datos para la prueba triangular

En la figura debajo se muestran los resultados arrojados por la función “prueba_binomial()”:

```

Número de respuestas correctas: 31
Para un p-valor de 0.00372349 , se rechaza la hipótesis nula y se concluye que
existe una diferencia significativa entre los productos evaluados.

Exact binomial test

data: exitos and n
number of successes = 31, number of trials = 60, p-value =
0.003723
alternative hypothesis: true probability of success is not equal to 0.3333333
95 percent confidence interval:
 0.3839460 0.6476871
sample estimates:
probability of success
      0.5166667

```

Figura 7.2 - Resultados de la prueba triangular

Se seguiría el mismo proceso para las pruebas tetrad, 2-AFC, 3-AFC, dúo-trío y dos de cinco.

7.2 Prueba DFC

Para ejemplificar el uso de la función prueba “prueba_DFC()”, se simularon datos correspondientes a una prueba DFC. Las puntuaciones relacionadas a la muestra de control oculto se generaron mediante una distribución $U[0,12]$ para representar resultados válidos donde los jueces identifican que el control oculto es similar a la referencia, mientras que las muestras bajo evaluación se simularon mediante una distribución $U[0,10]$ para cubrir todo el rango de la escala.

Debajo se muestran los primeros 6 valores (“head”) del dataframe que contiene el diseño experimental y las respuestas simuladas:

	Juez	Muestra1	Muestra2	id_Muestra1	id_Muestra2	muestra_A	resultados
1	1	RefA	B	664	718	No A	No A
2	2	RefA	B	296	526	No A	No A
3	3	RefA	A	720	196	A	No A
4	4	RefA	A	314	598	A	No A
5	5	RefA	B	145	690	No A	No A
6	6	RefA	B	322	278	No A	A

Figura 7.3 - Resultados simulación de datos para la prueba DFC

En la figura debajo se muestran los resultados arrojados por la función “prueba_DFC()” para el caso de los datos simulados, donde se ejecutó una prueba de Friedman debido a que no se cumplía la asunción de normalidad (a esperarse ya que se simularon mediante una distribución uniforme) y se están evaluando más de dos muestras.

Para un p-valor de 0.03315565 , se rechaza la hipótesis nula y se concluye que existe una diferencia significativa entre al menos uno de los productos evaluados y el control

Friedman rank sum test

data: Puntuacion and Producto and Juez
Friedman chi-squared = 6.8131, df = 2, p-value = 0.03316

Figura 7.4 - Resultados de la prueba DFC

7.3 Prueba “A” – “No A” / “Igual”-“Diferente”

Considerando que la diferencia entre las pruebas “A”-“No A” y la prueba “Igual”-“Diferente” radica en los aspectos cognitivos de la interpretación dada por los jueces, tanto la simulación de los datos como la ejecución de las pruebas es exactamente igual. Por tanto, se consolidará en un solo ejemplo, utilizando la prueba “A” – “No A” en su diseño monádico y pareado.

Debajo se muestran los primeros 6 valores (“head”) del dataframe que contiene el diseño experimental y las respuestas simuladas para el diseño monádico:

	Juez	Muestra1	Muestra2	id_Muestra1	id_Muestra2	muestra_A	resultados
1	1	RefA	B	664	718	No	No
2	2	RefA	B	296	526	No	No
3	3	RefA	A	720	196	Si	No
4	4	RefA	A	314	598	Si	No
5	5	RefA	B	145	690	No	No
6	6	RefA	B	322	278	No	Si

Figura 7.5 - Resultados simulación de datos para la prueba “A”-“No A” diseño monádico

La columna “muestra_A” identifica si la muestra presentada al juez realmente era una muestra A o no, esto para posteriormente compararla con la respuesta provista en la columna resultados.

Antes de proceder con el uso de estos datos, se convierte a una tabla de contingencia 2x2 utilizando las columnas “muestra_A” y “resultados”. En la figura debajo se muestran los resultados arrojados por la función “prueba_AnoA()” especificando el argumento diseño como “monadico”:

Para un valor p de 0.5622962 , no existe evidencia para determinar diferencia significativa entre los productos evaluados.

Pearson's Chi-squared test with Yates' continuity correction

data: datos
X-squared = 0.33574, df = 1, p-value = 0.5623

Figura 7.6 - Resultados de la prueba “A” – “No A” - diseño monádico

Debajo se muestran los primeros 6 valores (“head”) del dataframe que contiene el diseño experimental y las respuestas simuladas para el diseño pareado:

	Juez <int>	Muestra1 <fctr>	Muestra2 <fctr>	Muestra3 <fctr>	id_Muestra1 <int>	id_Muestra2 <int>	id_Muestra3 <int>	muestra_A1 <chr>	muestra_A2 <chr>	resultados_1 <chr>	resultados_2 <chr>
1	1	RefA	A	B	664	718	960	A	No A	A	No A
2	2	RefA	A	B	296	526	403	A	No A	A	No A
3	3	RefA	B	A	720	196	332	No A	A	A	No A
4	4	RefA	B	A	314	598	223	No A	A	No A	No A
5	5	RefA	A	B	145	690	650	A	No A	No A	No A
6	6	RefA	A	B	322	278	811	A	No A	No A	A

Figura 7.7 - Resultados simulación de datos para la prueba "A"- "No A" diseño pareado

Al igual que en el diseño anterior, la columna “muestra_A1” y “muestra A2” identifica si la muestra presentada al juez en primer y segundo orden realmente eran una muestra A o no, esto para posteriormente compararla con la respuesta provista en la columna resultados_1 y resultados_2 para el orden de presentación 1 y 2, respectivamente. Los datos se convierten a una tabla de contingencia 2x2 para identificar la distribución de las respuestas. En la figura debajo se muestran los resultados arrojados por la función “prueba_AnoA()” especificando el argumento diseño como “pareado”:

```
Para un valor p de 0.8743671 , no existe evidencia para determinar diferencia
significativa entre los productos evaluados.
```

```
McNemar's Chi-squared test with continuity correction
```

```
data: datos
```

```
McNemar's chi-squared = 0.025, df = 1, p-value = 0.8744
```

Figura 7.8 - Resultados de la prueba "A" - "No A" - diseño pareado

Conclusiones

Este trabajo fin de máster resume las distintas pruebas discriminativas que se utilizan en el análisis sensorial, destacando sus ventajas y desventajas de acuerdo con las características de los productos y los recursos limitantes que puedan presentar los clientes de Tastelab.

Por un lado, se revisaron los fundamentos teóricos de estas pruebas, incluyendo los aspectos cognitivos y la teoría estadística detrás de cada una ellas, lo que nos garantiza una comprensión amplia de las pruebas discriminativas. Por otro lado, se enfatizan requerimientos prácticos para la implementación de dichas pruebas, lo que nos proporciona la capacidad de analizar y entender los resultados sensoriales tomando en consideración todas las aristas del análisis sensorial.

Al examinar el análisis de la potencia estadística, se pretende que Tastelab pueda realizar estos análisis para valores que no estén incluidos en los estándares de análisis sensorial y que pueda tomar decisiones informadas en cuanto a la definición de la cantidad de jueces a participar en el estudio. También se propone seguir alimentando este análisis a medida que se hagan disponible más datos provenientes de las pruebas discriminativas implementadas.

Este trabajo fin de máster se podría considerar como una guía para el uso de pruebas discriminativas para el análisis sensorial mediante el lenguaje de programación R. La intención es que una vez se seleccione la prueba más adecuada en términos de las consideraciones cognitivas y recursos limitantes, su aplicación y análisis sea amigable para el usuario. Para esto se desarrollaron funciones a prueba de error que garantizaran la verificación de las asunciones de cada prueba y la retroalimentación al usuario para asegurar que los datos se están introduciendo correctamente.

Aunque las conclusiones se absan en pruebas de hipótesis y la determinación del p-valor, no quedan a interpretación del usuario ya que este trabajo fin de máster proporciona una base para que la conclusión de cada prueba discriminativa esté realmente asociada a las técnicas estadísticas que las soportan.

Bibliografía

- Meilgaard, M. C., Civille, G. V., & Carr, B. (2016). *Sensory Evaluation Techniques*. Florida: Taylor & Francis Group, LLC.
- Aberson, C. L. (2019). *Applied power analysis for the behavioral sciences* (2da edición ed.). Routledge. doi:978-1-315-17150-0
- Agresti, A. (2007). *An introduction to categorical data analysis* (Second edition ed.). Wiley-Interscience.
- American Society for Testing and Materials. (2016). Standard Test Method for Sensory Analysis - Tetrad Test (ASTM E3009-15e1). doi:10.1520/E3009-15E01
- Asociación Española de Normalización y Certificación. (2007). Análisis sensorial. Metodología. Guía general para la medición del olor, de la sensación olfátogustativa y del gusto mediante el procedimiento de elección forzosa de una entre tres alternativas (UNE-ISO 13301:2007).
- Asociación Española de Normalización y Certificación. (2007). Análisis sensorial. Metodología. Prueba de comparación por parejas. (UNE-EN ISO 5495:2007).
- Asociación Española de Normalización y Certificación. (2010). Análisis sensorial. Vocabulario. (UNE-EN ISO 5492:2010).
- Asociación Española de Normalización y Certificación. (2019). Análisis sensorial. Metodología. Guía general (UNE-ISO 6658:2019).
- Asociación Española de Normalización y Certificación. (2022). Análisis sensorial. Metodología. Prueba triangular. (UNE-EN ISO 4120:2022).
- Asociación Española de Normalización y Certificación. (2022). Análisis sensorial. Metodología. Prueba triangular. (UNE-EN ISO 4120:2022).
- Bausell, R. B., & Li, Y.-F. (2002). *Power analysis for experimental research: a practical guide for the biological, medical and social sciences*. Cambridge University Press.
- Bi, J. (2015). *Sensory Discrimination Tests and Measurements* (Second ed.). Wiley Blackwell. doi:9781118733530
- Bi, J., & Ennis, D. M. (1999). Statistical models for the A-Not A method. *The Institute for Perception*.
- Blitzstein, J. K., & Hwang, J. (2015). *Introduction to probability*. CRC Press. doi:978-1-4665-7559-2
- Bower, J. A. (2013). *Statistical Methods for Food Science*. Wiley Blackwell.
- Champely, S. (Oct de 2022). *pwr: Basic Functions for Power Analysis*. Obtenido de <https://cran.r-project.org/web/packages/pwr/index.html>
- Cho, H.-C., & Abe, S. (2012). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research* 66 (2013) 1261–1266.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second edition ed.). Lawrence Erlbaum Associates. doi:ISBN 0-8058-0283-5
- Dattalo, P. (2008). *Determining sample size*. Oxford University Press. doi:978-0-19-531549-3
- Dixon, W. F., & Massey, F. J. (1957). *Introduction to statistical analysis* (2nd edition ed.). McGraw-Hill.
- Kilcast, D. (2010). *Sensory analysis for food and beverage quality control*. Woodhead publishing limited.
- Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices* (Second ed.). Springer. doi:10.1007/978-1-4419-6488-5
- Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2013). *Introduction to probability & statistics* (Fourteenth ed.). Brooks / Cole Cengage Learning.
- Meyners, M. (2012). Equivalence tests – A review. *Food Quality and Preference*.
- Muñoz, A. M., Civille, G. V., & Carr, B. T. (1992). *Sensory evaluation in quality control*. Springer Science + Business Media LLC. doi:ISBN 978-1-4899-2655-5
- O'Mahony, M. (1986). *Sensory Evaluation of Food. Statistical methods and procedures*. Marcel Dekker Inc.
- Owen, D. B. (1965). The power of Student's t-test. *Journal of the American Statistical Association*.
- R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria. Obtenido de <https://www.R-project.org/>
- Rogers, L. (2017). *Discimination Testing in Sensory Science*. Woodhead publishing. doi:978-0-08-101116-4
- Stone, H., & Sidel, J. L. (2004). *Sensory Evaluation Practices* (Third ed.). Elsevier Academic Press.
- Triola, M. F. (2018). *Estadística* (Decimosegunda ed.). Pearson Educación de México. doi:978-607-32-4378-0
- Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications* (Seventh ed.). Thomson Brooks/Cole. doi:978-0-495-38508-0
- Yates, F. (1934). Contingency Tables Involving Small Numbers and the χ^2 Test. *Supplement to the Journal of the Royal Statistical Society, Vol. 1, No. 2 (1934), pp.217-235*.

Anexos

Anexo 1: Análisis de la potencia

Análisis de la potencia mediante fórmulas estadísticas

```
#Cargar librería para el análisis de la potencia
library("pwr")

#Definición de parámetros para las pruebas estadísticas

# Definición de parámetros para las pruebas estadísticas
tests <- c("binomial","t_student","ANOVA","chi_square") #pruebas a evaluar
alpha <- c(0.05, 0.01, 0.10) #niveles de significancia convencionales
efecto <- c(0.10,0.20,0.28) #magnitud del efecto
potencia <- seq(from=0.80, to=0.95, by=0.05) #potencia estadística

#Creación de combinación de parámetros
df <- expand.grid(test=tests, efecto=efecto, potencia=potencia,alpha=alpha)

#Creación de vector para almacenar el tamaño muestral
muestra <- numeric(nrow(df))

#Cálculo de tamaño muestral para cada combinación de parámetros
for (i in 1:nrow(df)) {
  if (df$test[i] == "binomial") {
    muestra[i] <- pwr.p.test(h=df$efecto[i], sig.level=df$alpha[i], power=df$potencia[i], alte
rnative="greater")$n
  } else if (df$test[i] == "t_student") {
    muestra[i] <- pwr.t.test(d=df$efecto[i], sig.level=df$alpha[i], power=df$potencia[i], alte
rnative="two.sided", n = NULL)$n
  } else if (df$test[i] == "ANOVA") {
    muestra[i] <- pwr.anova.test(f=df$efecto[i], sig.level=df$alpha[i], power=df$potencia[i],
k = 3)$n
  } else if (df$test[i] == "chi_square") {
    muestra[i] <- pwr.chisq.test(w=df$efecto[i], df=1, sig.level=df$alpha[i], power=df$potenci
a[i])$N
  }
}

#Agregar tamaño muestral al dataframe de los parámetros
df$n <- ceiling(muestra)

#Reorganización e impresión del dataframe
library(dplyr) #librería para manipulación de datos
df <- df %>% arrange(test) #reorganización basado en la prueba
print(format(df, scientific=FALSE))
```

Análisis de la potencia mediante simulación estadística

Prueba Binomial

```
#Definición de parámetros para la simulación
isim <- 10000 # Número de simulaciones
nsize <- seq(25, 150, by=10) # Secuencia de tamaño muestral
alpha <- c(0.01, 0.05, 0.10) # niveles de significancia

#Dataframe para almacenar p-valores
pv.bin <- data.frame(matrix(nrow = isim, ncol = length(nsize)))
colnames(pv.bin) <- as.character(nsize)

#Dataframe para almacenar proporción de rechazos
r.prop.bin <- data.frame(matrix(0, nrow = length(nsize), ncol = length(alpha)))
colnames(r.prop.bin) <- as.character(alpha)
rownames(r.prop.bin) <- as.character(nsize)

#Simulación

set.seed(2424) #Establecer semilla de simulación

for (j in 1:length(nsize)){
  for (i in 1:isim) {
    exitos <- rbinom(n=nsize[j],1,prob=0.5)
    resultado_prueba <- binom.test(x = sum(exitos), n = nsize[j], p=0.3)
    pv.bin[i,j] <- resultado_prueba$p.value
  }
  r.prop.bin[j,1] <- mean(pv.bin[,j]<0.01)
  r.prop.bin[j,2] <- mean(pv.bin[,j]<0.05)
  r.prop.bin[j,3] <- mean(pv.bin[,j]<0.10)
}

#Proporción de rechazos
colnames(r.prop.bin)=c("0.01", "0.05", "0.10")
print(r.prop.bin)

#Representación gráfica de la proporción de rechazos

#alpha 0.01
plot(nsize,r.prop.bin$`0.01`,type="b", ylim=c(0.2,1), ylab="Proporción de rechazo", xlab="Tamaño muestral",main="Análisis de potencia - Binomial")
points(nsize,r.prop.bin$`0.01`,col="red", pch="*")
lines(nsize,r.prop.bin$`0.01`,col="red",lty=1)
legend("bottomright",legend=c("0.01","0.05","0.10"), col=c("red","blue","green"), pch=c("o","*","*"),lty=c(1,1,1), ncol=1, title="Alpha")

#alpha 0.05
points(nsize,r.prop.bin$`0.05`,col="blue", pch="*")
lines(nsize,r.prop.bin$`0.05`,col="blue",lty=1)

#alpha 0.10
```

```
points(nsize,r.prop.bin$`0.10`,col="green", pch="*")
lines(nsize,r.prop.bin$`0.10`,col="green",lty=1)
```

Prueba T de Student

```
#Definición de parámetros para la simulación
isim <- 10000 # Número de simulaciones
nsize <- seq(25, 150, by=10) # Secuencia de tamaño muestral
alpha <- c(0.01, 0.05, 0.10) # niveles de significancia

#Dataframe para almacenar p-valores
pv.ts <- data.frame(matrix(nrow = isim, ncol = length(nsize)))
colnames(pv.ts) <- as.character(nsize)

#Dataframe para almacenar proporción de rechazos
r.prop.ts <- data.frame(matrix(0, nrow = length(nsize), ncol = length(alpha)))
colnames(r.prop.ts) <- as.character(alpha)
rownames(r.prop.ts) <- as.character(nsize)

#Simulación

set.seed(2424) #Establecer semilla de simulación

for (j in 1:length(nsize)){
  for (i in 1:isim) {
    x <- sample(c(0:2),size=nsize[j], replace=TRUE) #Data para la muestra control
    y <- sample(c(0:3),size=nsize[j], replace=TRUE) #Data para la muestra en evaluación
    pv.ts[i,j] <- t.test(x,y,"two.sided",paired=TRUE,mu=0)$p.value
  }
  r.prop.ts[j,1] <- mean(pv.ts[,j]<0.01)
  r.prop.ts[j,2] <- mean(pv.ts[,j]<0.05)
  r.prop.ts[j,3] <- mean(pv.ts[,j]<0.10)
}

#Proporción de rechazos
colnames(r.prop.ts)=c("0.01","0.05","0.10")
print(r.prop.ts)

#Representación gráfica de la proporción de rechazos

#alpha 0.01
plot(nsize,r.prop.ts$`0.01`,type="b", ylim=c(0,1), ylab="Proporción de rechazo", xlab="Tamaño
muestral",main="Análisis de potencia - T.Test")
points(nsize,r.prop.ts$`0.01`,col="red", pch="*")
lines(nsize,r.prop.ts$`0.01`,col="red",lty=1)
legend("bottomright",legend=c("0.01","0.05","0.10"), col=c("red","blue","green"), pch=c("o","*"),lty=c(1,1,1), ncol=1, title="Alpha")

#alpha 0.05
points(nsize,r.prop.ts$`0.05`,col="blue", pch="*")
lines(nsize,r.prop.ts$`0.05`,col="blue",lty=1)

#alpha 0.10
points(nsize,r.prop.ts$`0.10`,col="green", pch="*")
lines(nsize,r.prop.ts$`0.10`,col="green",lty=1)
```

Prueba ANOVA

```
#Definición de parámetros para la simulación
isim <- 10000 # Número de simulaciones
nsize <- seq(25, 150, by=10) # Secuencia de tamaño muestral
alpha <- c(0.01, 0.05, 0.10) # niveles de significancia

#Dataframe para almacenar p-valores
pv.aov <- data.frame(matrix(nrow = isim, ncol = length(nsize)))
colnames(pv.aov) <- as.character(nsize)

#Dataframe para almacenar proporción de rechazos
r.prop.aov <- data.frame(matrix(0, nrow = length(nsize), ncol = length(alpha)))
colnames(r.prop.aov) <- as.character(alpha)
rownames(r.prop.aov) <- as.character(nsize)

#Simulación

set.seed(2424) #Establecer semilla de simulación

for (j in 1:length(nsize)){
  for (i in 1:isim) {
    control <- sample(c(0:2),size=nsize[j], replace=TRUE) #Data para la muestra control
    ev1 <- sample(c(0:2),size=nsize[j], replace=TRUE) #Data para la muestra en evaluación
    ev2 <- sample(c(0:3),size=nsize[j], replace=TRUE) #Data para la muestra en evaluación
    data <- data.frame(puntuacion = c(control, ev1, ev2),
                      grupo = rep(c("control", "ev1", "ev2"), each = nsize[j]))
    aov_result <- aov(puntuacion ~ grupo, data = data)
    pv.aov[i,j] <- summary(aov_result)[[1]][["Pr(>F)"]][1]
  }
  r.prop.aov[j,1] <- mean(pv.aov[,j]<0.01)
  r.prop.aov[j,2] <- mean(pv.aov[,j]<0.05)
  r.prop.aov[j,3] <- mean(pv.aov[,j]<0.10)
}

#Proporción de rechazos
colnames(r.prop.aov)=c("0.01", "0.05", "0.10")
print(r.prop.aov)

#Representación gráfica de la proporción de rechazos

#alpha 0.01
plot(nsize,r.prop.aov$`0.01`,type="b", ylim=c(0,1), ylab="Proporción de rechazo", xlab="Tamaño
muestral",main="Análisis de potencia - ANOVA")
points(nsize,r.prop.aov$`0.01`,col="red", pch="*")
lines(nsize,r.prop.aov$`0.01`,col="red",lty=1)
legend("bottomright",legend=c("0.01", "0.05", "0.10"), col=c("red", "blue", "green"),
pch=c("o", "*"),lty=c(1,1,1), ncol=1, title="Alpha")

#alpha 0.05
points(nsize,r.prop.aov$`0.05`,col="blue", pch="*")
```



```

lines(nsize,r.prop.aov$`0.05`,col="blue",lty=1)

#alpha 0.10
points(nsize,r.prop.aov$`0.10`,col="green", pch="*")
lines(nsize,r.prop.aov$`0.10

```

Prueba de Wilcoxon

```

#Definición de parámetros para la simulación
isim <- 10000 # Número de simulaciones
nsize <- seq(25, 150, by=10) # Secuencia de tamaño muestral
alpha <- c(0.01, 0.05, 0.10) # niveles de significancia

#Dataframe para almacenar p-valores
pv.w <- data.frame(matrix(nrow = isim, ncol = length(nsize)))
colnames(pv.w) <- as.character(nsize)

#Dataframe para almacenar proporción de rechazos
r.prop.w <- data.frame(matrix(0, nrow = length(nsize), ncol = length(alpha)))
colnames(r.prop.w) <- as.character(alpha)
rownames(r.prop.w) <- as.character(nsize)

#Simulación

set.seed(2424) #Establecer semilla de simulación

for (j in 1:length(nsize)){
  for (i in 1:isim) {
    x <- sample(c(0:2),size=nsize[j], replace=TRUE) #Data para la muestra control
    y <- sample(c(0:3),size=nsize[j], replace=TRUE) #Data para la muestra en evaluación
    pv.w[i,j] <- suppressWarnings(wilcox.test(x,y,"two.sided",paired=TRUE)$p.value)
  }
  r.prop.w[j,1] <- mean(pv.w[,j]<0.01)
  r.prop.w[j,2] <- mean(pv.w[,j]<0.05)
  r.prop.w[j,3] <- mean(pv.w[,j]<0.10)
}

```

Prueba de Friedman

```

#Definición de parámetros para la simulación
diff <- 0.20 # Diferencia media entre los pares de observaciones
isim <- 5000 # Número de simulaciones
nsize <- seq(25, 150, by=10) # Secuencia de tamaño muestral
alpha <- c(0.01, 0.05, 0.10) # niveles de significancia

#Dataframe para almacenar p-valores
pv.f <- data.frame(matrix(nrow=isim, ncol=13))
names <- c("n=25","n=35","n=45","n=55","n=65","n=75","n=85","n=95","n=105","n=115","n=125","n=135","n=145")
colnames(pv.f) <- names

#Dataframe para almacenar proporción de rechazos
r.prop.f <- data.frame(row.names=c("n=25","n=35","n=45","n=55","n=65","n=75","n=85","n=95","n=105","n=115","n=125","n=135","n=145"))

```

```

#Simulación

set.seed(2424) #Establecer semilla de simulación

for (j in 1:length(nsize)){
  for (i in 1:isim) {
    #Data para la muestra control
    control <- runif(nsize[j],0,2)
    #Data para la primera muestra en evaluación
    ev1 <- runif(nsize[j],0,3)
    #Data para la segunda muestra en evaluación
    ev2 <- runif(nsize[j],0,4)

    #Matriz para la data de las muestras
    data <- as.matrix(data.frame(control,ev1,ev2))

    pv.f[i,j] <- friedman.test(data)$p.value
  }
  r.prop.f[j,1] <- mean(pv.f[,j]<0.01)
  r.prop.f[j,2] <- mean(pv.f[,j]<0.05)
  r.prop.f[j,3] <- mean(pv.f[,j]<0.10)
}

#Proporción de rechazos
colnames(r.prop.f)=c("0.01","0.05","0.10")
print(r.prop.f)

#Representación gráfica de la proporción de rechazos

#alpha 0.01
plot(nsize,r.prop.f$`0.01`,type="b", ylim=c(0,1), ylab="Proporción de rechazo", xlab="Tamaño m
uestral",main="Análisis de potencia - Friedman")
points(nsize,r.prop.f$`0.01`,col="red", pch="*")
lines(nsize,r.prop.f$`0.01`,col="red",lty=1)
legend("bottomright",legend=c("0.01","0.05","0.10"), col=c("red","blue","green"), pch=c("o","*
"),lty=c(1,1,1), ncol=1, title="Alpha")

#alpha 0.05
points(nsize,r.prop.f$`0.05`,col="blue", pch="*")
lines(nsize,r.prop.f$`0.05`,col="blue",lty=1)

#alpha 0.10
points(nsize,r.prop.f$`0.10`,col="green", pch="*")
lines(nsize,r.prop.f$`0.10`,col="green",lty=1)

```

Anexo 2: Diseño experimental

Posible órdenes de las muestras

```

# Prueba triangular
triangular_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"))
triangular_samples <- triangular_samples[!(rowSums(triangular_samples == "B") == 3 | rowSums(
triangular_samples == "A") == 3), ]
triangular_samples

#Prueba tetrad
tetrad_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"),c("A", "B"))
tetrad_samples <- tetrad_samples[rowSums(tetrad_samples == "A") == 2 & rowSums(tetrad_samples
== "B") == 2,]
rename_with(tetrad_samples, ~sub("Var", "Muestra", .))

```

```

tetrad_samples

#Prueba dos de cinco
tof_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"), c("A", "B"), c("A", "B"))
tof_samples <- tof_samples[(rowSums(tof_samples == "A") == 2 & rowSums(tof_samples == "B") ==
3) |
                        (rowSums(tof_samples == "A") == 3 & rowSums(tof_samples == "B") ==
2), ]
rename_with(tof_samples, ~sub("Var", "Muestra", .))
tof_samples

#Prueba 2-AFC
AFC2_samples <- expand.grid(c("A", "B"), c("A", "B"))
AFC2_samples <- AFC2_samples[(rowSums(AFC2_samples == "B") == 1 & rowSums(AFC2_samples == "A"
) == 1), ]
rename_with(AFC2_samples, ~sub("Var", "Muestra", .))
AFC2_samples

#Prueba 3-AFC
AFC3_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"))
AFC3_samples <- AFC3_samples[!(rowSums(AFC3_samples == "B") == 3 | rowSums(AFC3_samples == "A"
) == 3), ]
rename_with(AFC3_samples, ~sub("Var", "Muestra", .))
AFC3_samples

#Prueba duo-trio
duotrio_samples <- expand.grid(c("RefA", "ControlA", "B"), c("RefA", "A", "B"), c("RefA", "A",
"B"))

duotrio_samples <- duotrio_samples[duotrio_samples[,1] == "RefA" &
((duotrio_samples[,2] == "A"& duotrio_samples[,3] == "B") |
(duotrio_samples[,2] == "B" & duotrio_samples[,3] == "A")), ]

rename_with(duotrio_samples, ~sub("Var", "Muestra", .))
duotrio_samples

#Prueba diferente del control
DFC_samples <- expand.grid(c("A", "B", "C"), c("A", "B", "C"), c("A", "B", "C"))

DFC_samples <- DFC_samples[
                        rowSums(DFC_samples[,] == "A") == 1 &
                        rowSums(DFC_samples[,] == "B") == 1 &
                        rowSums(DFC_samples[,] == "C") == 1, ]
rename_with(DFC_samples, ~sub("Var", "Muestra", .))
DFC_samples

#Prueba A-No A: Diseño monádico
AnoA_monadico_samples <- expand.grid(c("RefA", "A"), c("A", "B"))
AnoA_monadico_samples <- AnoA_monadico_samples[AnoA_monadico_samples[,1] == "RefA", ]
rename_with(AnoA_monadico_samples, ~sub("Var", "Muestra", .))
AnoA_monadico_samples

#Prueba A-No A: Diseño pareado
AnoA_pareado_samples <- expand.grid(c("RefA", "A", "B"), c("RefA", "A", "B"), c("RefA", "A", "B"
))
AnoA_pareado_samples <- AnoA_pareado_samples [AnoA_pareado_samples[,1] == "RefA" &
                        rowSums(AnoA_pareado_samples[,] == "A") == 1 &
                        rowSums(AnoA_pareado
_samples[,] == "B") == 1, ]
rename_with(AnoA_pareado_samples, ~sub("Var", "Muestra", .))

```

```
AnoA_pareado_samples
```

```
#Prueba Igual-Diferente: Diseño monádico
```

```
ID_monadico_samples <- expand.grid(c("RefA", "A"), c("A", "B"))  
ID_monadico_samples <- ID_monadico_samples[ID_monadico_samples[,1] == "RefA", ]  
rename_with(ID_monadico_samples, ~sub("Var", "Muestra", .))  
ID_monadico_samples
```

```
#Prueba Igual-Diferente: Diseño pareado
```

```
ID_pareado_samples <- expand.grid(c("RefA", "A", "B"), c("RefA", "A", "B"), c("RefA", "A", "B"))  
ID_pareado_samples <- ID_pareado_samples [ID_pareado_samples[,1] == "RefA" & rowSums(ID_pareado_samples[,] == "A") == 1 &  
rowSums(ID_pareado_samples[,] == "B") == 1,  
]  
rename_with(ID_pareado_samples, ~sub("Var", "Muestra", .))  
ID_pareado_samples
```

Asignación de las muestras

```
#dplyr
```

```
library(dplyr)
```

```
# Función para generar el diseño experimental
```

```
generar_diseño <- function(prueba, num_jueces)
```

```
{  
  design <- switch (prueba, "triangular"= {  
    #Definición de las posibles presentaciones de las muestras  
    triangular_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"))  
    triangular_samples <- triangular_samples[!(rowSums(triangular_samples == "B") == 3 | rowSums(triangular_samples == "A") == 3), ]  
  }
```

```
    # Aleatorización de la presentación de las muestras a cada juez
```

```
    triangular_samples %>% sample_n(num_jueces, replace=TRUE) %>%
```

```
    mutate(Juez = 1:num_jueces) %>%
```

```
    rename_with(~sub("Var", "Muestra", .))
```

```
  }, "tetrad"= {
```

```
    #Definición de las posibles presentaciones de las muestras
```

```
    tetrad_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"), c("A", "B"))
```

```
tetrad_samples <- tetrad_samples[rowSums(tetrad_samples == "A") == 2 & rowSums(tetrad_samples == "B") == 2, ]
```

```
rename_with(tetrad_samples, ~sub("Var", "Muestra", .))
```

```
    # Aleatorización de la presentación de las muestras a cada juez
```

```
    tetrad_samples %>% sample_n(num_jueces, replace=TRUE) %>%
```

```
    mutate(Juez = 1:num_jueces) %>%
```

```
    rename_with(~sub("Var", "Muestra", .))
```

```
  }, "2of5"= {
```

```
    #Definición de las posibles presentaciones de las muestras
```

```
    tof_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"), c("A", "B"), c("A", "B"))
```

```
tof_samples <- tof_samples[(rowSums(tof_samples == "A") == 2 & rowSums(tof_samples == "B") == 3) |
```

```
(rowSums(tof_samples == "A") == 3 & rowSums(tof_samples == "B") == 2), ]
```

```
rename_with(tof_samples, ~sub("Var", "Muestra", .))
```

```
    # Aleatorización de la presentación de las muestras a cada juez
```

```
    tof_samples %>% sample_n(num_jueces, replace=TRUE) %>%
```

```
    mutate(Juez = 1:num_jueces) %>%
```

```
    rename_with(~sub("Var", "Muestra", .))
```

```
  }, "2AFC"= {
```

```
    #Definición de las posibles presentaciones de las muestras
```

```
AFC2_samples <- expand.grid(c("A", "B"), c("A", "B"))
```

```

AFC2_samples <- AFC2_samples[(rowSums(AFC2_samples == "B") == 1 & rowSums(AFC2_samples == "A"
) == 1), ]
rename_with(AFC2_samples, ~sub("Var", "Muestra", .))

# Aleatorización de la presentación de las muestras a cada juez
AFC2_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "3AFC"= {
#Definición de las posibles presentaciones de las muestras
AFC3_samples <- expand.grid(c("A", "B"), c("A", "B"), c("A", "B"))
AFC3_samples <- AFC3_samples[!(rowSums(AFC3_samples == "B") == 3 | rowSums(AFC3_samples == "A"
) == 3), ]
rename_with(AFC3_samples, ~sub("Var", "Muestra", .))

# Aleatorización de la presentación de las muestras a cada juez
AFC3_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "duo_trio"= {
#Definición de las posibles presentaciones de las muestras
duotrio_samples <- expand.grid(c("RefA", "ControlA", "B"), c("RefA", "A", "B"), c("RefA", "A",
"B"))

duotrio_samples <- duotrio_samples[duotrio_samples[,1] == "RefA" &
((duotrio_samples[,2] == "A" & duotrio_samples[,3] == "B") |
(duotrio_samples[,2] == "B" & duotrio_samples[,3] == "A")), ]

rename_with(duotrio_samples, ~sub("Var", "Muestra", .))

# Aleatorización de la presentación de las muestras a cada juez
duotrio_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "DFC"= {
#Definición de las posibles presentaciones de las muestras
DFC_samples <- expand.grid(c("A", "B", "C"),c("A", "B", "C"),c("A", "B", "C"))

DFC_samples <- DFC_samples[
rowSums(DFC_samples[,] == "A") == 1 &
rowSums(DFC_samples[,] == "B") == 1 &
rowSums(DFC_samples[,] == "C") == 1, ]
rename_with(DFC_samples, ~sub("Var", "Muestra", .))
# Aleatorización de la presentación de las muestras a cada juez
DFC_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "AnoA_monadico"= {
#Definición de las posibles presentaciones de las muestras
AnoA_monadico_samples <- expand.grid(c("RefA", "A"), c("A", "B"))
AnoA_monadico_samples <- AnoA_monadico_samples[AnoA_monadico_samples[,1] == "RefA", ]
rename_with(AnoA_monadico_samples, ~sub("Var", "Muestra", .))
# Aleatorización de la presentación de las muestras a cada juez
AnoA_monadico_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "AnoA_pareado"= {
#Definición de las posibles presentaciones de las muestras
AnoA_pareado_samples <- expand.grid(c("RefA", "A", "B"),c("RefA", "A", "B"),c("RefA", "A", "B"
))

```

```

AnoA_pareado_samples <- AnoA_pareado_samples [AnoA_pareado_samples[,1] == "RefA" &
  rowSums(AnoA_pareado_samples[,] == "A") == 1 &
  rowSums(AnoA_pareado
_samples[,] == "B") == 1, ]
rename_with(AnoA_pareado_samples, ~sub("Var", "Muestra", .))

# Aleatorización de la presentación de las muestras a cada juez
AnoA_pareado_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "ID_monadico"= {
#Definición de las posibles presentaciones de las muestras
ID_monadico_samples <- expand.grid(c("RefA", "A"), c("A", "B"))
ID_monadico_samples <- ID_monadico_samples[ID_monadico_samples[,1] == "RefA", ]
rename_with(ID_monadico_samples, ~sub("Var", "Muestra", .))

# Aleatorización de la presentación de las muestras a cada juez
ID_monadico_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
}, "ID_pareado"= {
#Definición de las posibles presentaciones de las muestras
ID_pareado_samples <- expand.grid(c("RefA", "A", "B"),c("RefA", "A", "B"),c("RefA", "A
", "B"))
ID_pareado_samples <- ID_pareado_samples [ID_pareado_samples[,1] == "RefA" & rowSums(ID_parea
do_samples[,] == "A") == 1 &
rowSums(ID_pareado_samples[,] == "B") == 1,
]
rename_with(ID_pareado_samples, ~sub("Var", "Muestra", .))

# Aleatorización de la presentación de las muestras a cada juez
ID_pareado_samples %>% sample_n(num_jueces,replace=TRUE) %>%
mutate(Juez = 1:num_jueces) %>%
rename_with(~sub("Var", "Muestra", .))
},
stop("Prueba no reconocida. Elegir entre triangular, tetrad, 2of5, 2AFC, 3AFC, duo_trio, D
FC, AnoA_monadico, AnoA_pareado,ID_monadico ó ID_pareado")
)

#Identificadores de la muestras
design <- design %>%
mutate(across(starts_with("Muestra"), ~ sample(100:999, n(), replace = FALSE), .names = "id_
{.col}"))

#Reordenación de las columnas
patron_muestra <- "Muestra[0-9] +"
patron_id <- "Id[0-9] +"
columnas <- c("Juez", grep(patron_muestra, names(design), value = TRUE), grep(patron_id
, names(design), value = TRUE))
design <- select(design, all_of(columnas))

#Impresión del resultado
return(design)
}

# Uso de la función para generar el diseño experimental
set.seed(12424)
diseño <- generar_diseño("AnoA_pareado", 5)
diseño

#

```

Anexo 3: Implementación de las técnicas estadísticas

Prueba binomial

```
# Función para la ejecución prueba binomial
prueba_binomial <- function(prueba, exitos, n, alpha=0.05) {

  p <- switch(prueba,
    "triangular" = 1/3,
    "tetrad_esp" = 1/6,
    "tetrad_no" = 1/3,
    "2of5" = 1/10,
    "2AFC" = 1/2,
    "3AFC" = 1/3,
    "duo_trio" = 1/2,
    stop("Prueba no reconocida. Elegir entre triangular, tetrad_esp, tetrad_no, 2of5
, 2AFC, 3AFC, duo_trio, DFC, AnoA_monadico, AnoA_pareado, ID_monadico ó ID_pareado")
  )
  resultado_prueba <- binom.test(x = exitos, n = n, p = p)

  # Impresión número de respuestas correctas
  cat("Número de respuestas correctas:", exitos, "\n")

  # Imprimir el resultado de la prueba
  if (resultado_prueba$p.value < alpha) {
    cat("Para un p-valor de", resultado_prueba$p.value, ", se rechaza la hipótesis nula y se co
ncluye que existe una diferencia significativa entre los productos evaluados.", "\n")
  } else {
    cat("Para un p-valor de", resultado_prueba$p.value, ", no existe evidencia para determinar
diferencia significativa entre los productos evaluados.", "\n")
  }

  return(resultado_prueba)
}
```

Prueba DFC

```
# Función para la ejecución prueba diferente de control
prueba_DFC <- function(resultados, n_productos, alpha=0.05) {
  #comprobación de normalidad para prueba t-test y ANOVA
  p_values <- apply(resultados[,paste0("Puntuacion", 1:n_productos)], 2, function(x) shapiro.t
est(x)$p.value)
  normalidad <- all(p_values > 0.05)

  #Selección de la prueba a realizar
  if(n_productos == 2) {
    # Si se evalúan dos productos, se realiza una prueba t de student o Wilcoxon
    if(normalidad) {
      # Se realiza prueba t de Student si los datos son normales
      test_results <- t.test(resultados$Puntuacion1, resultados$Puntuacion2)
    } else {
      # Se realiza prueba de Wilcoxon si los datos no son normales
      test_results <- wilcox.test(resultados$Puntuacion1, resultados$Puntuacion2)
    }
  } else {
    # Si se evalúan más de dos productos, se realiza un ANOVA o prueba de Friedman
    if(normalidad) {
      #Identificación de la fórmula para aplicación de la función
      formula_string <- paste("Puntuacion1 ~ ", paste(paste0("Puntuacion", 2:n_productos), collapse=
" + "))
    }
  }
}
```

```

    # Se realiza un ANOVA si Los datos son normales
    test_results <- aov(as.formula(formula_string), data = resultados)
  } else {
    #Reestructuración de Los resultados
    resultados_long <- resultados %>%
      pivot_longer(starts_with("Puntuacion"), names_to = "Producto", values_to = "Puntuacion")
    test_results <- friedman.test(Puntuacion ~ Producto | Juez, data = resultados_long)
  }
}

# Imprimir conclusión
if (test_results$p.value < alpha) {
  cat("Para un p-valor de", test_results$p.value, ", se rechaza la hipótesis nula y se concluye que existe una diferencia significativa entre al menos uno de los productos evaluados y el control", "\n")
} else {
  cat("Para un p-valor de", test_results$p.value, ", no existe evidencia para determinar diferencia significativa entre los productos evaluados.", "\n")
}

#Imprimir todo el resultado
return(test_results)
}

```

Prueba chi cuadrado: “A” - “No A”

```

prueba_chisqA <- function(datos, diseño, alpha = 0.05) {

  # Verificaciones iniviales
  if (any(dim(datos) != 2)) {
    stop("Los datos tienen que estar en una tabla de contingencia 2x2")
  }

  # Verificación de uso corrección de Yates
  frecuencias_esperadas <- sum(datos) * prod(dim(datos)) / prod(sum(datos))
  correccion <- any(frecuencias_esperadas < 5)
  #Realización de prueba de acuerdo al diseño de la prueba
  if (diseño == "monadico") {
    # Pearson Chi-Square de Homogeneidad
    test_results <- chisq.test(datos, correct = correccion)
  } else if (diseño == "mixto") {
    # Pearson Chi-Square de Independencia
    test_results <- chisq.test(datos, correct = correccion)
  } else if (diseño == "pareado") {
    # McNemar Chi-Square de Proporción Correlacionada
    test_results <- mcnemar.test(datos, correct = correccion)
  } else {
    stop("Diseño no reconocido. Elegir entre monadico, mixto o pareado.")
  }

  # Imprimir conclusión
  if (diseño == "monadico") {
    if(test_results$p.value < alpha) {
      cat("Para un valor p de", test_results$p.value, ", se rechaza la hipótesis nula y se concluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que responden 'A' no es homogénea entre la presentación de las muestras 'A' y 'No A'. \n")
    } else {
      cat("Para un valor p de", test_results$p.value, ", no existe evidencia para determinar diferencia significativa entre los productos evaluados.\n")
    }
  }
}

```



```

} else if (diseño == "pareado") {
  if(test_results$p.value < alpha) {
    cat("Para un valor p de", test_results$p.value, ", se rechaza la hipótesis nula y se con
cluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que
responden 'A' está relacionado con la presentación de las muestras 'A' y 'No A'.\\n")
  } else {
    cat("Para un valor p de", test_results$p.value, ", no existe evidencia para determinar d
iferencia significativa entre los productos evaluados.\\n")
  }
} else {
  if(test_results$p.value < alpha) {
    cat("Para un valor p de", test_results$p.value, ",se rechaza la hipótesis nula y se conc
luye que existe una diferencia significativa entre los productos ya que la muestra presentada
tiene un efecto significativo sobre la respuesta 'A' o 'No A'.\\n")
  } else {
    cat("Para un valor p de", test_results$p.value, ", no existe evidencia para determinar d
iferencia significativa entre los productos evaluados.\\n")
  }
}

return(test_results)
}

```

Prueba chi cuadrado: “Igual” - “Diferente”

```

prueba_chisqID <- function(datos, diseño, alpha = 0.05) {

  # Verificaciones iniviales
  if (any(dim(datos) != 2)) {
    stop("Los datos tienen que estar en una tabla de contingencia 2x2")
  }

  # Verificación de uso corrección de Yates
  frecuencias_esperadas <- sum(datos) * prod(dim(datos)) / prod(sum(datos))
  correccion <- any(frecuencias_esperadas < 5)
  #Realización de prueba de acuerdo al diseño de la prueba
  if (diseño == "monadico") {
    # Pearson Chi-Square de Homogeneidad
    test_results <- chisq.test(datos, correct = correccion)
  } else if (diseño == "mixto") {
    # Pearson Chi-Square de Independencia
    test_results <- chisq.test(datos, correct = correccion)
  } else if (diseño == "pareado") {
    # McNemar Chi-Square de Proporción Correlacionada
    test_results <- mcnemar.test(datos, correct = correccion)
  } else {
    stop("Diseño no reconocido. Elegir entre monadico, mixto o pareado.")
  }

  # Imprimir conclusión
  if (diseño == "monadico") {
    if(test_results$p.value < alpha) {
      cat("Para un valor p de", test_results$p.value, ", se rechaza la hipótesis nula y se con
cluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que
responden 'Igual' no es homogénea entre la presentación de las muestras 'Igual' y 'Diferente'
.\\n")
    } else {
      cat("Para un valor p de", test_results$p.value, ", no existe evidencia para determinar d
iferencia significativa entre los productos evaluados.\\n")
    }
  } else if (diseño == "pareado") {
    if(test_results$p.value < alpha) {

```

```

    cat("Para un valor p de", test_results$p.value, ", se rechaza la hipótesis nula y se con
cluye que existe diferencia significativa entre los productos ya que la cantidad de jueces que
responden 'Igual' está relacionado con la presentación de las muestras 'Igual' y 'Diferente'.
\n")
  } else {
    cat("Para un valor p de", test_results$p.value, ", no existe evidencia para determinar d
iferencia significativa entre los productos evaluados.\n")
  }
} else {
  if(test_results$p.value < alpha) {
    cat("Para un valor p de", test_results$p.value, ",se rechaza la hipótesis nula y se conc
luye que existe una diferencia significativa entre los productos ya que la muestra presentada
tiene un efecto significativo sobre la respuesta 'Igual' o 'Diferente'. \n")
  } else {
    cat("Para un valor p de", test_results$p.value, ", no existe evidencia para determinar d
iferencia significativa entre los productos evaluados.\n")
  }
}
}

return(test_results)
}

```

Anexo 4: Simulación de los datos y ejecución de las pruebas

Prueba binomial: datos prueba triangular

```

library(dplyr)
set.seed(12424) #semilla para asegurar reproducibilidad
diseño <- generar_diseño("triangular", 60) #generación del diseño experimental
#Identificación de respuesta correcta
diseño <- diseño %>% mutate(respuesta_correcta = if_else(Muestra1 != Muestra2 & Muestra1 !=
Muestra3 , id_Muestra1,
  if_else(Muestra2 != Muestra1 & Muestra2 != Muestra3 , id_Muestra2,
    if_else(Muestra3 != Muestra1 & Muestra3 != Muestra2 , id_Muestra3,NA_integer_))))

#Generación de los datos: simulación de la respuesta de los jueces
set.seed(12424) #semilla para asegurar reproducibilidad
respuesta_juez <- apply(diseño[, c("id_Muestra1", "id_Muestra2", "respuesta_correcta")], 1, fu
nction(x) sample(x, 1, prob=c(0.3,0.3,0.3)))

#Determinación de la cantidad de respuestas correctas (exitos)
resultados <- cbind(diseño,respuesta_juez)
resultados <- resultados%>% mutate(identificacion_correcta = if_else(respuesta_juez == respues
ta_correcta, "Correcta", "Incorrecta"))

exitos <- sum(resultados$identificacion_correcta == "Correcta")

#Determinacion de la cantidad de ensayos
n <- nrow(resultados)

#Ejecución de la prueba binomial
prueba_binomial ("triangular",exitos,n,alpha=0.05)

```

Prueba DFC

```

library(dplyr)
library(tidyverse)
n_jueces <- 60 #numero de jueces
set.seed(12424) #semilla para asegurar reproducibilidad
diseño <- generar_diseño("DFC", n_jueces) #generación del diseño experimental
#Identificación de respuesta correcta
diseño <- diseño %>% mutate(control_oculto = if_else(Muestra1=="A" , id_Muestra1,

```

```

    if_else(Muestra2 == "A" , id_Muestra2,
    if_else(Muestra3 == "A", id_Muestra3, NA_integer_)))

#Generación de los datos: simulación de la respuesta de los jueces (magnitud de la diferencia)
set.seed(12424) #semilla para asegurar reproducibilidad

#Resultados de los jueces
resultados <- diseño %>%
  mutate(
    Puntuacion1 = case_when(
      Muestra1 == "A" ~ sample(c(0:2),size=n_jueces, replace=TRUE),
      Muestra1 == "B" ~ sample(c(0:8),size=n_jueces, replace=TRUE),
      Muestra1 == "C" ~ sample(c(0:10),size=n_jueces, replace=TRUE)
    ),
    Puntuacion2 = case_when(
      Muestra2 == "A" ~ sample(c(0:2),size=n_jueces, replace=TRUE),
      Muestra2 == "B" ~ sample(c(0:8),size=n_jueces, replace=TRUE),
      Muestra2 == "C" ~ sample(c(0:10),size=n_jueces, replace=TRUE)
    ),
    Puntuacion3 = case_when(
      Muestra3 == "A" ~ sample(c(0:2),size=n_jueces, replace=TRUE),
      Muestra3 == "B" ~ sample(c(0:8),size=n_jueces, replace=TRUE),
      Muestra3 == "C" ~ sample(c(0:10),size=n_jueces, replace=TRUE)
    )
  )

#Ejecución de la prueba DFC
prueba_DFC (resultados, n_productos=3,alpha=0.05)

```

Prueba "A no A": Diseño monádico

```

library(tidyverse)
n_jueces <- 60 #numero de jueces
set.seed(12424) #semilla para asegurar reproducibilidad
diseño <- generar_diseño("AnoA_monadico", n_jueces) #generación del diseño experimental
#Identificación de respuesta correcta
diseño <- diseño %>% mutate(muestra_A = if_else(Muestra2=="A" , "A", "No A",))

#Generación de resultados
resultados <- diseño %>% mutate(resultados = sample(c("A","No A"),n(),replace=TRUE,prob=c(0.3,0.7)))

#Conversión a tabla de contingencia
resultados <- table(resultados$muestra_A, resultados$resultados)

#Ejecución de la prueba DFC
prueba_chisqA (resultados, diseño="monadico",alpha=0.05)

```

Prueba "A no A": Diseño pareado

```

library(tidyverse)
n_jueces <- 60 #numero de jueces
set.seed(12424) #semilla para asegurar reproducibilidad
diseño <- generar_diseño("AnoA_pareado", n_jueces) #generación del diseño experimental
#Identificación de respuesta correcta
diseño <- diseño %>% mutate(muestra_A1 = if_else(Muestra2=="A" , "A", "No A",)) %>% mutate(muestra_A2 = if_else(Muestra3=="A" , "A", "No A",))

#Generación de resultados
resultados <- diseño %>% mutate(resultados_1 = sample(c("A","No A"),n(),replace=TRUE,prob=c(

```

```

0.3,0.7)), resultados_2 = sample(c("A", "No A"), n(), replace=TRUE, prob=c(0.3,0.7))

head(resultados)
#Conversión a tabla de contingencia

resultados <- resultados %>%
  mutate(respuesta_1 = if_else(muestra_A1 == resultados_1, "Correcto", "Incorrecto"),
         respuesta_2 = if_else(muestra_A2 == resultados_2, "Correcto", "Incorrecto"))
tabla <- table(resultados$respuesta_1, resultados$respuesta_2)

#Ejecución de La prueba DFC
prueba_chisqA (tabla, diseño="pareado", alpha=0.05)

```