UNIVERSIDAD SANTO TOMÁS AQUINO

División de Ciencias Económicas y administrativas

Facultad de estadística

Herramientas para depuración y análisis de datos En investigación de mercados

Trabajo de grado para optar por el título de estadístico

Estudiante:
JEFERSON ALEJANDRO GUEVARA NUÑEZ

Director:
JOSE FERNANDO ZEA CASTRO

Bogotá D.C, Colombia

Septiembre de 2015

CONTENIDO

CONTENIDO DE TABLAS

CONTENIDO DE CUADROS

CONTENIDO DE GRÁFICOS

CONTENIDO DE ILUSTRACIONES

INTRODUCCIÓN

En el presente trabajo de grado se busca la elaboración de herramientas computacionales para llevar a cabo la depuración y análisis de información proveniente de estudios de investigación de mercados. Los desarrollos realizados en este programa son de fácil acceso a un público general puesto que se desarrollan con herramientas de software libre y permiten evitar el uso de software licenciado.

La primera parte del documento contiene una introducción a la investigación de mercado para tener un panorama general de los procesos que se llevan a cabo usualmente en una amplia gama de estudios de mercadeo, se resumen los conceptos básicos de mercadeo que permiten hacer uso de las herramientas computacionales para llevar a cabo la depuración de la información recolectada en los estudios y realizar algunos análisis cuantitativos ampliamente utilizados en la investigaciones de mercados.

Posteriormente se da una visión amplia de los problemas que se presentan con la información recolectada y se presentan algunas soluciones para llevar a cabo la depuración de los conjuntos de datos que se utilizan en los estudios de mercados, lo anterior es vital ya que de la depuración de la información depende que los resultados que se obtienen tengan validez y permitan una adecuada toma de decisiones.

Los problemas abordados para la depuración de información provenientes de estudios de mercados son algunos de los más frecuentes que se presentan con las tablas de datos y que conllevan dificultades en su procesamiento y análisis. Estos son: inconsistencia de manejo de respuestas múltiples, problemas de coherencia entre

conjuntos de variables múltiples, duplicidad de información, y problemas en las validaciones de datos en los análisis de sensibilidad de precios.

Se llevaron a cabo desarrollos computacionales en software estadístico para las siguientes técnicas de análisis de investigación de mercados: Modelo Kano, cálculo del Índice de Jaccard, Análisis de Penalidades (conocido en la literatura anglosajona como penalty analysis), y el análisis de sensibilidad de Precios de Van Westendorp (en inglés Van Westendorp's Price Sensitivity Meter - PSM). Existen muchas otras técnicas utilizadas en investigación de mercado, la escogencia de estas técnicas obedecen a su popularidad y facilidad de interpretación.

El trabajo desarrollado en esta tesis tiene un alcance práctico en el ámbito profesional y no solamente académico ya que está pensado para que los técnicos y profesionales que laboran en el campo de la investigación de mercados lo puedan utilizar.

Los datos que se presentan en este trabajo para ilustrar el uso del software desarrollado son simulados, pero están basados en datos reales utilizados en agencias de investigación de mercados.

Para cada función de depuración y análisis desarrollada se llevará a cabo una descripción de la técnica, elaboración de la función en R, documentación de la función, y desarrollo de ejemplos de la función.

JUSTIFICACIÓN

En cualquier estudio que involucre recolección de datos y en particular en los estudios de mercados se presentan problemas con la consistencia de la información. Algunos de los problemas que ocurren con los datos surgen desde la construcción del formulario de recolección en el cual en ocasiones no se dispone de controles de validación y de consistencia adecuados, otros problemas surgen por errores humanos que se cometen con la herramienta de captura. Se hace necesario por lo tanto realizar una revisión a las bases de datos para verificar que la información cuenta con los elementos necesarios para posteriores análisis, por tal razón se agruparon los problemas más frecuentes que se pueden presentar y servir como ayuda al personal encargado del procesamiento de la información como también a estadísticos, ya que para cualquier análisis es necesario que la base cumpla con unas especificaciones establecidas.

El presente trabajo busca ilustrar los problemas más usuales que se presentan en bases de datos provenientes de estudios de mercadeo y dar solución a ellos a través de rutinas computacionales en software libre altamente flexibles que reduzcan tiempos y sean de fácil utilización.

Por otro lado se desarrollan programas para implementar algunos de los métodos de análisis más usados en investigación de mercados tales como el análisis de penalidades, el análisis de sensibilidad de precios entre otros. Varios de estos análisis no se encuentran disponibles en ningún software estadístico de uso general o se encuentran en software privativo los cuales requieren de una licencia de pago. Los desarrollos de los

métodos en R no solamente permitirán una amplia difusión de los métodos sino que permitirá un acceso a un público amplio de personas que laboran en el ámbito de la investigación de mercados.

OBJETIVOS

Objetivo general

- Elaborar herramientas computacionales para llevar a cabo la depuración y análisis de información proveniente de estudios de mercadeo.

Objetivos específicos

- Elaborar un estado del arte de los conceptos básicos de mercadeo desde un punto de vista cuantitativo y presentar algunos de los análisis usuales utilizados en los estudios de mercadeo.

- Mostrar los problemas típicos que se presentan con la información estadística utilizada para llevar a cabo un estudio de investigación de mercados y elaborar rutinas en software estadístico que permita su solución.

- Elaborar rutinas de análisis en software estadístico que permitan implementar los análisis usuales en investigación de mercados.

MARCO TEÓRICO

1. GENERALIDADES

1.1. Introducción breve a la investigación de mercadeo

Un estudio de mercado puede tener varios significados, según (Sarstedt & Mooi, 2014) puede ser el proceso por el cual se obtiene una idea de cómo funcionan los mercados, una función en una organización o puede hacer referencia a los resultados de la investigación, como una base de datos de compras de los clientes o de un informe con recomendaciones.

Algunas personas consideran que la investigación de mercados y estudios de mercado son sinónimo, mientras que otros consideran estos como conceptos diferentes. La Asociación Americana de Mercadeo (*American Marketing Association*), la mayor asociación de marketing en Estados Unidos, define la investigación de mercados de la siguiente manera:

“La función que vincula al consumidor, cliente y público para el vendedor a través de información que se utiliza para identificar los problemas, oportunidades de negocio, además de generar, refinar y evaluar acciones, monitorear el desempeño de la comercialización”.

En el proceso de investigación de mercados se diseña el método de acopio de información, se gestiona y ejecuta el proceso de recopilación de datos se analizan los

resultados y se comunican los hallazgos y sus implicaciones (American Marketing Association, 2004)

Por otro lado, ESOMAR la organización mundial para el mercado, los consumidores y la investigación social, define la investigación de mercado como:

“La recopilación sistemática y la interpretación de la información sobre las personas o las organizaciones que utilizan los métodos y técnicas estadísticas y analíticas de las ciencias sociales aplicadas para obtener conocimientos o apoyar la toma de decisiones. La identidad de los encuestados no se dará a conocer al usuario de la información sin el consentimiento explícito y no se hará ningún enfoque de ventas para ellos como resultado directo de haber proporcionado información (ICC / ESOMAR International Code on Market and Social, 2007).

Ambos definiciones se incorporan sustancialmente pero la definición de la AMA se centra en la investigación de mercados como una función (por ejemplo, un departamento en una organización), mientras que la definición de ESOMAR se centra en el proceso.

Para comprender mejor la dinámica de la investigación de mercados (Sarstedt & Mooi, 2014) desarrollan un ejemplo con el automovil *Prius de Toyota*, un automóvil híbrido que utiliza gasolina / motor eléctrico.

Existían algunos antecedentes para desarrollar coches *frugales* (ahorrativos, eléctricos) por parte de Honda y de General Motors que no habían funcionado bien.

Después de resolver diferentes problemas técnicos relacionados con el sistema de integración de la gasolina y el motor eléctrico, en 1999, Toyota tomó la decisión de

empezar a trabajar en el lanzamiento del *Prius* en los Estados Unidos, la investigación inicial de mercado demostró que iba a ser una difícil tarea, algunos consumidores pensaban que era demasiado pequeño para los Estados Unidos y algunos pensaron que el posicionamiento de los controles fue pobre para los conductores estadounidenses. Había otros temas también, como el diseño, que muchos pensaban era demasiado fuerte y dirigido a los conductores japoneses.

Mientras se preparaba para el lanzamiento, Toyota llevó a cabo más estudios de mercado, lo que podría, sin embargo, no revelaban que compradores potenciales del tendría el automóvil. En un principio, Toyota cree que el automóvil podría ser comprarlo las personas interesadas en el medio ambiente, pero la investigación de mercado descartó esta hipótesis, otros estudios de mercados hicieron poco para identificar cualquier otro segmento de mercado. A pesar de la falta de hallazgos concluyentes, Toyota decidió vender el automóvil de todos modos y para esperar las reacciones del público. Antes del lanzamiento, Toyota puso un sistema de investigación de mercado en el lugar para realizar un seguimiento de las ventas iniciales e identificar donde los clientes compraron el automóvil. Después del lanzamiento formal en 2000, este sistema rápidamente encontró que el automóvil estaba siendo comprado por celebridades para demostrar su preocupación por el medio ambiente. Algo más tarde, Toyota notó considerablemente aumentó las cifras de ventas cuando los consumidores habituales se dieron cuenta de la demanda del auto por las celebridades. Parecía que los consumidores están dispuestos a comprar los coches avalados a causa de las celebridades.

CNW Market Research, una empresa de investigación de mercado especializada en la industria del automóvil, que atribuye parte del éxito del *Prius* a su diseño único, que

demostraron claramente que los propietarios de *Prius* estaban conduciendo un automóvil diferente. Después de un aumento sustancial en el precio de la gasolina, y los cambios en el automóvil (sobre la base de una amplia investigación de mercado) para aumentar su atractivo, Toyota alcanzó unas ventas totales de más de tres millones de dólares y ahora es el líder del mercado de automóviles híbridos gasolina / eléctricos.

Este ejemplo muestra que, si bien la investigación de mercado de vez en cuando ayuda, a veces contribuye poco o incluso falla. Hay muchas razones por las que el éxito de la investigación de mercado varía. Estas razones incluyen el presupuesto disponible para la investigación, el apoyo a la investigación de mercados en la organización, ejecución y las habilidades de investigación de los investigadores de mercado

.

Las fases de la investigación de mercados en muchas de las organizaciones, es el siguiente:

1.2. Tipos de Investigación de mercados

Investigación Cualitativa: se utiliza para conocer los efectos que causa un estímulo a los entrevistados, tiene una parte más personal, donde el moderador (puede ser un psicólogo) indaga profundamente sobre aquellos efectos mencionados, también averigua si las pruebas de venta o los beneficios del producto o servicio ofrecido y la manera en que se informan son creíbles.

Investigación Cuantitativa: busca cuantificar los datos, utiliza la recolección de datos y análisis para probar hipótesis establecidas previamente y contestar peguntas con respecto a la investigación.

Otros tipos de investigación pero menos frecuentes se encuentra la Investigación Documental: se basa de fuentes de carácter documental, esto es, en documentos de cualquier género.

1.3. Etapas de una Investigación de mercado

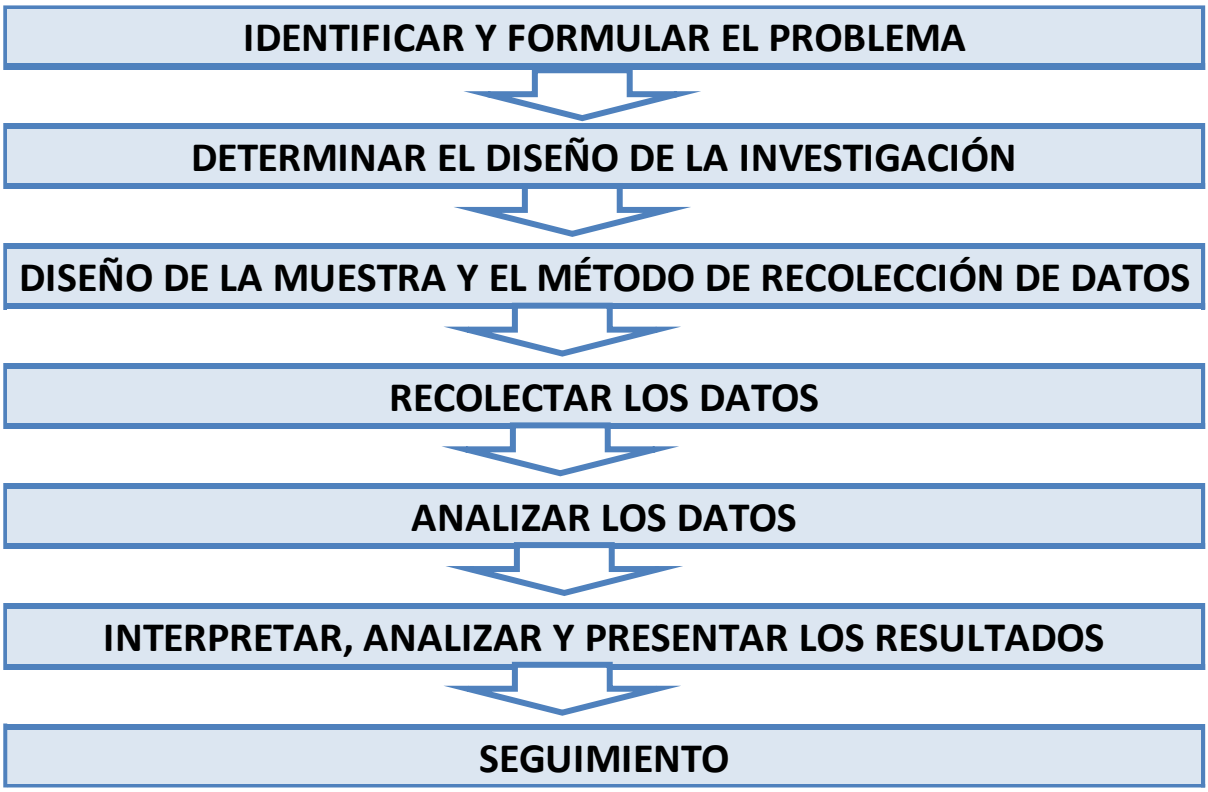


Ilustración 1 Mapa de Procesos en una Invetigación (Sarstedt & Mooi, 2014)

a. Identificar y formular el problema

Este paso es el muy importante, ya que identificar el problema de investigación es valioso, pero también difícil, ya que se deben tener en cuenta variable como el mercado (Ver Palabras Claves), examinar los síntomas de mercadeo o las oportunidades de mercadeo, tales como la disminución de participación de una marca en el mercado, aumento de número de quejas sobre un producto o

compañía, porque el producto nuevo lanzado al mercado no tiene éxito, entre otros.

b. **Determinar el diseño de la Investigación**

El diseño de la investigación se relaciona con la identificación y formulación del problema. Los problemas de investigación y diseños de investigación son altamente relacionada. Si se comienza a trabajar en un tema que nunca se ha investigado antes, parece que entrar en un embudo donde inicialmente pedimos preguntas exploratorias ya que aún no se sabe acerca de los problemas a los que se enfrentan. Estas preguntas exploratorias se responden mejor utilizando un diseño de investigación exploratoria. Una vez que se tenga una idea más clara de la asunto de investigación y después de la investigación exploratoria, se puede describir el problema de la investigación en términos de investigación descriptiva (Ver Palabras Claves). Una vez se tenga una imagen razonablemente completa de todos los temas, puede ser el momento para determinar exactamente cómo se vinculan las variables clave. Por lo tanto, se pasa a la parte más estrecha del embudo es decir es como una depuración. Cada diseño de investigación tiene diferentes usos y requiere la aplicación de diferentes técnicas de análisis. Por ejemplo, mientras que la investigación exploratoria puede ayudar a formular problemas exactamente o estructurarlos, investigación causal proporciona información exacta sobre cómo las variables se relacionan (Sarstedt & Mooi, 2014).

c. Diseño de la muestra y el método de recolección de datos

Luego de determinar el diseño de la investigación, se debe diseñar un plan de muestreo y elegir un método de recolección de datos. Se trata de decidir si utilizar los datos existentes o para llevar a cabo la investigación nueva.

Entre los métodos de recolección que más se utilizan, puede encontrarse: encuestas o formularios en papel, en tabletas, celulares o computadores con programas diseñados para este fin, encuestas grabadas (las cuales conllevan una transcripción posterior).

d. Recolectar los datos

La recolección de la información es una práctica pero a veces difícil ya que debe tenerse en cuenta como se diseña una encuesta, también como medir los atributos hacia un producto, marca o empresa. Hacer frente a problemas requiere una planificación y el conocimiento del proceso ya que de omitirse alguna información valiosa puede acarrear análisis defectuosos, perdida de dinero.

e. Analizar los datos

El análisis los datos requiere de habilidades técnicas. Debe tenerse en cuenta que los datos deben estar limpios (tener una depuración adecuada), en el trabajo de grado se especifica algunos de las revisiones comunes que permiten identificar falencias en la información, además se explican cuatro análisis que tienen

participación en la investigación de mercados, y que sirven como ayuda a soluciones comunes para productos o servicios.

f. Interpretar, analizar y presentar los resultados

Cuando se ejecuta el proceso de investigación de mercado, uno de los objetivos inmediatos está interpretando, discutiendo y presentando los hallazgos. En consecuencia, los investigadores deben dar respuestas detalladas y sugerencias de acciones concretas basadas en los datos y técnicas de análisis (Sarstedt & Mooi, 2014).

g. Seguimiento

Las agencias investigación de mercado y los investigadores de mercadeo a menudo se detienen cuando los resultados han sido interpretados, discutido y presentado. Sin embargo, el seguimiento de los hallazgos de investigación también es importante. La implementación de los hallazgos de investigación de mercado a veces requiere más indagación, porque las sugerencias o recomendaciones pueden ya no sean la mismas y las condiciones del mercado pueden haber cambiado. La investigación de seguimiento en la investigación realizada previamente puede ser una buena manera de entrar en nuevos acuerdos para llevar a cabo más investigaciones.

Algunos estudios de mercado no se acaban nunca. Por ejemplo, muchas empresas hacen seguimiento de la satisfacción del cliente con el tiempo. Incluso este tipo de investigación puede tener seguimiento, por ejemplo, la administración puede desear saber sobre caídas en la satisfacción del cliente.

1.4. El personal encargado de la depuración de las bases de datos tenemos los siguientes roles

Revisor: es el encargado de inspeccionar que el cuestionario se haya llevado a cabo de forma correcta, si la encuesta es en papel lo hará de forma manual

Procesador de datos: es la persona encargada de generar las tablas de resumen del estudio, también se encargan de validar la información por medio de programación y evitar posibles errores.

Estadístico: genera las muestras y los análisis estadísticos que se presenten es el desarrollo del estudio.

Analista: es el que tiene como propósito la función de interpretar la información obtenida de las tablas y/o análisis especiales que se presenten en el estudio.

En esta sección facilitaremos la labor del procesador de datos y del estadístico para obtener resultados de calidad que permitan realizar análisis confiables.

Las bases de datos que llegan para procesamiento en muchos casos presentan problemas que dificultan realizar los procedimientos estadísticos de análisis. Ignorar estos problemas conlleva a conclusiones erróneas, además de dificultar las tareas del personal encargado de procesamiento y del estadístico.

1.5. Tipos de estudio

Dependiendo la agencia cambia el nombre de los productos u ofrecen más, pero generalmente los que se realizan son:

- Prueba de producto – Pre-test: Encaminado a productos nuevos, es el primer vistazo del producto.
- Branding (Evaluaciones de marca): Evaluaciones que miden la marca o marcas propia y de la competencia, termómetro de cómo está la marca.
- Awareness (Conocimiento, percepción, recuerdo): Estudios realizados con el fin de ver la recordación por ejemplo un producto nuevo lanzado unos meses atrás, las preguntas están encaminadas en ver si fue impactante la publicidad en torno al lanzamiento del producto o marca, también podría servir como seguimiento.

Pensando en el seguimiento, existen estudios como:

- Panales: estos estudios reportan información periódica, donde las personas que forman parte de este panel, se comprometen a ser parte de este, durante un tiempo, el tiempo está determinado según las especificaciones explicitas del panel.
- Post-test: los estudios que miden el comportamiento de un producto después del pre-test, las preguntas están encaminadas en la evolución y en el desenvolvimiento del producto en el mercado.

2. PALABRAS CLAVES

Dentro del trabajo se trabajaran términos los cuales se deben ampliar para tener un concepto más extenso de la información que allí se hable, primero se puede mencionar cosas como:

La ***Encuesta,*** que desde el punto de vista de la investigación de mercados se desarrolla preguntándoles a los encuestados asuntos concretos. Tiene el propósito de conseguir información sobre actitudes, motivos y opiniones. Las ***Encuestas*** se pueden realizar por medio de visitas personales, por teléfono o por correo.

Dentro de las ***Encuestas*** podemos encontrar preguntas con ***Escala Likert*** se denomina así por Rensis Likert, que en 1932 publicó un informe donde describía su uso. Es una escala psicométrica usualmente utilizada en cuestionarios de investigación de mercados. Al responder a una pregunta de un cuestionario elaborado con la técnica de Likert, se detalla el nivel de acuerdo o desacuerdo con una declaración.

De estas escalas también podemos obtener ***Top Two Box (TTB –T2B)*** que corresponde a la suma de los dos porcentajes positivos obtenidos en una variable, de la misma manera el ***Top Tree Box*** que hace referencia a los tres porcentajes positivos obtenidos en una variable, ahora bien, el ***Top Box*** hace relación a la porcentaje positivo mayor de una escala.

Otro tipo de preguntas que usualmente se usan son ***Top of Mind*** que es el indicador que revela cuál es la ***Marca*** que, cuando le preguntan por una categoría específica, se le viene a la mente en primer lugar al mayor porcentaje de personas. También se le llama primera mención, porque claramente está de primera en la mente y es mencionada de forma espontánea en primer lugar.

Además aunque no es usual se calculan ***Promedios,*** dicho ***Promedio*** es la suma de todos los valores obtenidos al multiplicar el número de respondientes por la calificación que dieron; la suma total se divide entre la muestra y así se obtiene el promedio.

Las variables ***dummies*** son aquellas variables dicotómicas de tipo cualitativo, también las han llamado variables binarias, que pueden asumir los valores de 0 y 1, mostrando la presencia o ausencia de una cualidad o atributo.

Cuando se habla de ***Marca*** o ***Brand*** (***Brand*** *es* la ***Marca*** de producto o de un servicio) nos referimos a un nombre, término, signo, símbolo o diseño, o la combinación de todos ellos, que tiende a identificar bienes o servicios de un vendedor o grupo de vendedores y diferenciarlo de los la competencia en el **Mercado**.

El concepto de ***Mercado*** es muy amplio ya que la expresión se utiliza con frecuencia en el lenguaje común y para relatar distintas realidades, una de las definiciones referentes al concepto encaminada a la investigación sería, que es un grupo reconocible de consumidores con poder adquisitivo, que están dispuestos y disponibles para pagar por un producto o un servicio o también podría ser la totalidad de los compradores potenciales y actuales de algún producto o servicio.

Los ***Atributos de un producto*** que vienen siendo como la personalidad del mismo, tienen una serie de factores que permiten realizar un examen del producto, comenzando por los elementos propios de la marca hasta los complementarios, para que a la vista tanto los productos propios como de los de la competencia, para que se pueda elaborar la estrategia del mercadeo que permita posicionar un producto en el mercado de la manera más propicia.

Investigación Descriptiva es una forma de estudio de investigación, donde se busca saber dónde, quién, cómo, cuándo y por qué del sujeto del estudio, es decir un estudio descriptivo debe explicar completamente al entrevistado, objetos, servicios o cuentas.

Entre los estudios más realizados podemos encontrar ***Estudios Ad-hoc,*** que pueden tener un carácter cualitativo o cuantitativo. Son estudios no sistemáticos, realizados “a medida”. Existen tantos tipos de estudios ad- hoc como problemas de marketing puedan surgir en una empresa.

También encontramos estudios que miden el ***Rating***, que es el ***Rating*** es el porcentaje de la audiencia total que está viendo un programa o anuncio.

3. PROBLEMAS FRECUENTES DE BASES DE DATOS

La mayoría de departamentos y empresas de mercadeo tienen problemas con sus bases de datos, ya se a problemas de capturas si la información se encuentra en papel, ya sea porque sea una encuesta auto administrada o virtual, o bien, que provenga de información que el cliente proporciona; estos problemas conllevan a resultados no adecuados y por consiguiente análisis incorrectos, por tal razón y buscando dar soluciones a algunos de estos inconvenientes se enumerara seis problemas frecuentes que se presentan con la información almacenada en bases de datos:

3.1. Problemas reestructuración de bases con rotación de productos

Un problema usual que se presentan en las bases de datos que contienen información proveniente de investigaciones de mercado tiene que ver con la consistencia de los datos aquellas en donde se efectúan comparaciones entre dos o más productos y se efectúan rotaciones de los mismos para eliminar sesgo de elección por parte de los participantes.

Suponga que se requiere comparar dos fragancias en cuanto a sus atributos tales como el olor, la duración, entre otros. Para llevar a cabo esto a cada uno de los participantes se le presenta los diferentes productos y se desarrolla una encuesta en donde ellos evalúan cada uno de los atributos.

Suponga que se disponen de dos productos que compiten en el mercado, y se requiere comparar sus desempeño en diferentes atributos (A_1, A_2, …,A_p). Para reducir el sesgo de percepción se predefine de forma aleatoria el orden en el que se evalúan los productos, al anterior procedimiento se le denomina rotación.

Primera Evaluación	Segunda Evaluación
Producto 1	Producto 2
Producto 1	Producto 2
Producto 2	Producto 1
Producto 1	Producto 2
Producto 2	Producto 1
Producto 2	Producto 1
Producto 2	Producto 1
Producto 1	Producto 2

Tabla 1: Elección del orden de evaluación de los productos

Cada uno de los productos se les evalúa los mismos atributos en una escala *Likert* (Ver Palabras Claves) (por ejemplo de uno a cinco), sin embargo para capturar la calificación de cada uno ellos no necesariamente se utilizan las mismas columnas. En lugar de esto se almacena la información de los atributos de acuerdo al orden previamente definido para evaluar los productos, esto con el fin de eliminar sesgos, cabe mencionar que para este ejercicio se rotan solo los productos pero en otros procedimientos se rotan tanto productos como atributos.

Se ilustra lo anterior en la Tabla 2 (Problemas reestructuración de bases). Suponga que se escogen a cinco individuos para que comparen dos productos (producto 1 y producto 2).

Como se observa en la primera fila, los atributos del primer producto se evalúan al comienzo, luego se evalúan los atributos del segundo producto. La información de los cuatro atributos que se evalúan del producto es capturada en las columnas comprendidas entre P_1 y P_4. Por otra parte el segundo producto que se evalúa es el segundo (ver Alt2), se evalúan exactamente los mismos atributos y estos se capturan entre las columnas P_5 a P_8.

Por otra parte para el tercer individuo se le asigna para evaluar en primer lugar al producto dos, para el cual se le captura la información de los cuatro atributos en las columnas comprendidas entre P_1 y P_4. El segundo atributo que se evalúa es el producto dos para el cual la calificación de los atributos es recolectada en las columnas P_5 a la P_8.

El anterior esquema es fácilmente ampliado para tres o más productos. Es usual recolectar información de la preferencia global entre uno de los dos productos, en este caso se presenta dicha información en la columna (Preferencia).

ID	ALT1	P1	P2	P3	P4	ALT2	P5	P6	P7	P8	Preferencia
1	1	5	5	5	1	2	5	4	5	1	1
2	2	2	3	4	1	1	4	4	4	2	1
3	2	4	4	5	1	1	5	5	5	1	1
4	1	4	4	4	1	2	4	3	5	1	1
5	2	3	2	3	1	1	3	2	4	2	2
6	1	4	4	4	2	2	4	3	4	1	1

Tabla 2: Problemas reestructuración de bases

Para poder llevar a cabo el procesamiento de la base de datos se construyen programas en el software R que disponga la información del segundo producto escogido por cada uno de los individuos y lo disponga de la siguiente forma:

ID	ALT1	P1	P2	P3	P4	Preferencia
1	1	5	5	5	1	1
2	2	2	3	4	1	1
3	2	4	4	5	1	1
4	1	4	4	4	1	1
5	2	3	2	3	1	2
6	1	4	4	4	2	1
ID	ALT2	P5	P6	P7	P8	Preferencia
1	2	5	4	5	1	1
2	1	4	4	4	2	1
3	1	5	5	5	1	1
4	2	4	3	5	1	1
5	1	3	2	4	2	2
6	2	4	3	4	1	1

Tabla 3: Reestructuración de Base de Datos

Para obtener la solución final se lleva a cabo dos reestructuraciones clásicas de una base ancha (*wide format*) a una base larga (*long format*) a la tabla anterior. Obteniéndose la tabla siguiente:

ID	ALT	P1	P2	P3	P4	Preferencia
1	1	5	5	5	1	1
1	2	2	5	4	5	1
2	1	4	4	4	2	1
2	2	2	3	4	1	1
3	1	5	5	5	1	1
3	2	4	4	5	1	1
4	1	4	4	4	1	1
4	2	4	3	5	1	1
5	1	3	2	4	2	2
5	2	3	2	3	1	2
6	1	4	4	4	2	1
6	2	4	3	4	1	1

Tabla 4: Reestructuración de bases con rotación de productos

Con la información anterior en ocasiones se presentan problemas de coherencia en la escogencia de la preferencia del producto. Ilustraremos lo anterior con otro ejemplo; suponga que se evalúan tres atributos de dos productos diferentes que se recogen respectivamente para el primer producto en las columnas P_1 a P_3 y para el segundo producto en las columnas P_5 a P_7.

Por ejemplo, para los dos productos se podrían evaluar atributos como el olor, el color y la densidad y para uno de los dos productos obtenerse un mayor promedio de las calificaciones de los atributos pero marcarse como preferencia el otro producto, esto sería un problema de coherencia con la elección de la preferencia.

En la Tabla 5 (Comparación de Preferencia) se observa que el entrevistado de la quinta fila prefiere el producto dos sin embargo el promedio de los atributos del producto uno es 3 y el promedio del producto dos es 2.7, lo que se espera es que el producto la preferencia sea el producto 1 y no el producto 2 como lo marca el ejemplo.

P7	P8	Preferencia		Revisión 1 ALT1 <> ALT2	medio ALT1 y P		Comparación Promedio vs Preferencia
5	1	1	→	VERDADERO	5.0	4.7	VERDADERO
4	2	1		VERDADERO	3.0	4.0	VERDADERO
5	1	1		VERDADERO	4.3	5.0	VERDADERO
5	1	1		VERDADERO	4.0	4.0	VERDADERO
4	2	2		VERDADERO	2.7	3.0	FALSO
4	1	1		VERDADERO	4.0	3.7	VERDADERO

Tabla 5: Comparación de Preferencia

Sin embargo, la única solución es a este problema es verificar la información si la encuesta fue realizada por medio físico, o si se tiene evidencia de las respuestas en algun lugar.

3.2. Problemas Respuestas Múltiples

En las preguntas de respuesta múltiple no es posible colocar toda la información en una variable, así por ejemplo cuando se indaga por las diferentes marcas que consume un individuos se debe disponen en diferentes columnas la información. Como se mencionó anteriormente cada columna corresponde al orden de mención de las diferentes opciones y se plasman los códigos que identifiquen a cada uno de los diferentes productos en las diferentes columnas.

Para que la información resultante de una pregunta de respuestas múltiples sea coherente se debe verificar que un mismo individuo no tenga repetida una o más respuestas, por ejemplo suponga que en una pregunta se indaga a cada individuo cuales fueron las características que le agradaron más de un producto, supongamos que esa pregunta se codifica como P1; cada una de las respuestas asociadas a estas preguntas son P1.1 para la primera mención, P1.2 para la segunda y así sucesivamente según la cantidad de respuestas dadas cada uno, Suponga que previamente se identificaron 25 características que pueden gustar de un producto y que la persona 1 responde que prefiere la característica 21 (P1.1=21) que corresponde a el sabor y vuelve a mencionar en la cuarta mención la misma característica (P1.4=21), es decir el individuo responde dos veces la misma característica (el sabor), esto no implica que el sabor sea su mayor elección, generalmente estos problemas se deben a la confusión del encuestado o a

inconvenientes en la captura de la información y debe por tal razón identificarse y darle la solución establecida en el estudio que usualmente es quitar esta segunda mención del mismo código, ver Tabla 6 (Problemas Respuestas Múltiples).

ID	P1.1	P1.2	P1.3	P1.4	P1.5	P1.6
1	21	26	30	21		
2	12					
3	12					
4	1					
5	12					
6	12	21				

Tabla 6: Problemas Respuestas Múltiples

Para dar solución a este problema debe identificarse también los valores faltantes (por ejemplo códigos 99 = No sabe y 98 = No responde), estos códigos no deberían tenerse en cuenta al momento de comparar si existen duplicados.

El procedimiento para la identificación de estos casos duplicados mencionados en este ejemplo es primero tener un prefijo que generalmente es el comienzo de la pregunta, por ejemplo “P1.1, P1.2” en este caso P1 se denominaría como un prefijo que hace referencia a la pregunta, también debemos tener un separador que identifique donde se cambia un nombre de una variable de otra, de nuevo con el ejemplo “P1.1, P1.2” el “.” seria nuestro separador que nos identifica los datos que van antes del prefijo y después de él.

El programa que se desarrolla para identificar duplicados en respuestas múltiples indicará si las respuestas del individuo presentan o no problema de duplicados en la respuesta múltiple.

3.3. Problema de datos faltantes

La falta de información o valores faltantes ocurre frecuente en los datos provenientes de estudios de mercadeo, muchas veces el encuestador omite o diligencia mal el formulario o en otros casos el encuestado se rehúsa a dar información.

Describiremos a continuación un problema frecuente que ocurre en las bases de datos de los estudios de mercados, suponga que se le realiza al encuestado la siguiente pregunta ¿mencione cuatro marcas de gaseosas que conoce?.

Ahora se puede observar el caso de la Tabla 7 (Estructura Base Problema de datos faltantes), donde el ID es el número de la encuesta, en este caso el ID=1 se tiene tres respuestas para una pregunta en diferentes momentos (Marcas), para apreciar mejor las respuestas dadas se asignó un color a las marca, la pregunta realizada es con respuesta espontánea:

MARCA = ¿mencione cuatro marcas de gaseosas que conoce?, la persona podía contestar las marcas que se le vinieran a la cabeza es decir que máximo podía contestar cuatro pero podría no recordar sino dos o una, recuerde que los números consignados en la base de datos en MARCA1 a MARCA4 hacen referencia a marcas de gaseosas sin importar su posición es decir en MARCA1 es la primera marca que se le viene en mente, lo primero que se debe mirar es que la persona no me conteste la misma marca más de una vez como en el anterior referente a problema de respuestas múltiples; Las demás respuestas (P1 a P7) están dadas con respecto a esa Marca es decir P1.1 corresponde a la marca dada para la variable MARCA1, P1.2 corresponde a la marca dada para la variable MARCA2 y así sucesivamente. El color de la celda hace referencia a la marca:

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5	5	4	5	2		
2	1	2	3		4	4	2		4	2	2	
3	1	2	3	5	4	5	4	4	4	2	4	3
4	2	3	5		2	3	4		4	4	4	
5	1	2	5		3	3	1		4	1	3	
6	1	2	3		2	3	3		4	2	2	

Tabla 7: Estructura Base Problema de datos faltantes

El primer caso que se puede apreciar que para la pregunta P1. ¿Qué calificación de 1 a 5 donde uno es nada satisfecho y cinco muy satisfecho, le da usted a la marca … con respecto a su calidad?, en la celda P1.4 del ID = 1, muestra un valor de 4, el cual no debería existir ya que solo recordó 3 marcas, otro caso es en la misma encuesta en el ID = 1, la pregunta P2 ¿Cómo calificaría de 1 a 5 donde uno es muy agradable y 5 muy agradable el sabor de la gaseosa de la marca…?, en la celda P2.3 se encuentra un dato sin respuesta, en Tabla 8 (Falta de Información), en color rojo se encuentra los problemas mencionados:

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5	5	4	5	2		
2	1	2	3		4	4	2		4	2	2	
3	1	2	3	5	4	5	4	4	4	2	4	3
4	2	3	5		2	3	4		4	4	4	
5	1	2	5		3	3	1		4	1	3	
6	1	2	3		2	3	3		4	2	2	

Tabla 8: Falta de Información

Para dar solución a estos problemas se deben identificar primero las inconsistencias citadas, después se debe verificar con un medio físico si fuese el caso que la encuesta se realizara en papel, de lo contrario se debe buscar un método de

comprobación como por ejemplo recontactar el encuestado e indagar de nuevo, de no ser posible darle solución se debe anular la respuesta.

3.4. Problemas con Duplicidad de Información

En este problema se busca identificar si existe duplicidad total o parcial de la información, por ejemplo en el ejemplo Tabla 9 (Problemas con Duplicidad de Información) se observa en color rojo una encuesta duplicada en su totalidad, la cual tiene la misma información pregunta por pregunta, lo que se debe hacer en estos casos es verificar primero y determinar el problema de la duplicidad y proceder a eliminar una de las dos encuestas, pues puede ser que se presentara la inconsistencia desde la captura de la información, por lo cual debe quitarse una, ahora bien, en ejemplo también se observa una encuesta duplicada en solamente en la variable ID (Identificación) la cual se marca en rojo, ya que el ID es el identificador o llave de la encuesta no debe presentarse más de una vez a menos que la base de datos se encuentre estructurada de otra forma distinta, para un análisis especifico diferente, para dar solución a este problema primero se debe verificar que no contiene las dos encuestas la es la misma información (en color verde se muestra que el resto de la encuesta no es duplicidad), luego de confirmar que son distintas las encuestas se debe cambiar el número del ID por un número que no se duplique con ninguna otra encuesta capturada en la base de datos, la forma de asignar el número es comprobando si se la información se encuentra en papel y asignar el correspondiente, de no contar con un material en físico conviene estipular un método donde se el número que debe llevar, esto a criterio del dueño del proyecto.

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5	5		5	2	5	
2	1	2	3		4	4	2		4	2	2	
3	1	2	3	5	4	5	4	4	4	2	4	3
4	2	3	5		2	3	4		4	4	4	
5	1	2	5		2	2	2		4	4	3	
1	1	2	5		5	5	5		5	2	5	
6	1	3			4	4			4	2		
7	2	3	5		2	2	2		4	4	3	
8	1	2	5		3	2	3		4	5	4	
2	3	4			4	5			5	1		

Tabla 9: Problemas con Duplicidad de Información

3.5. Problemas de Coherencia

En el desarrollo de una encuesta en investigación de mercados existen preguntas que están relacionadas unas con otras, para ilustrar el problema el siguiente ejemplo donde P1 (P1.RU a P1.OMR12), P1.RU hace referencia en este caso a un *Top of Mind* (Ver Palabras Claves) donde la pregunta puede ser ¿cuándo hablamos de crema dental , cual es la primera marca que se le viene a la mente?, y P1.OMR1 a P1.OMR12 otras respuestas ¿Qué otras marcas de crema dental recuerda?; las respuestas son teniendo en cuenta las marcas relacionadas en P1 ya que dependen de si la mencionaron o no, en P2 ¿de las marcas que ya me menciono que recuerda cuales ha utilizado en el último año? menciono la marca 1 en P1 debió nombrarse también (ver Tabla 10 Problemas de Coherencia en color verde), de igual manera en la pregunta P3 ¿de las marcas que conoce cuales ha utilizado en el último mes?, se colocaron colores a las marcas en cada celda para identificar en la base donde se presentan estas inconsistencias, la solución para este problema es identificar los casos donde mencionaron la marca que no se encuentra en P1, ya que se debe verificar si la encuesta está en medio físico, de lo

contrario se debe contactar al encuestado, de no ser así conviene retirarse la marca que se mencionó en P2 pero que no está en P1.

ID	P1.RU	P1.OMR1	P1.OMR2	P1.OMR3	...	P1.OMR12	P2R1	P2R2	P2R3	P2R4	...	P2R12	P3R1	P3R2	P3R3	P3R4	...	P3R12
1000	6	1	2				2	3	6	714			2	3	6	705		
1002	6	1	2	3			1	3	6				1	6				
1003	650	1	3	6			1	6	650				1	650				
1004	3	1	6				1	3	6				3					
1006	3	5	6				3	5	6				3	5				

Tabla 10: Problemas de Coherencia

3.6. Problemas de Coherencia en Precios

Tiene que ver con un tema de Coherencia de los datos, en este caso los precios se toman de esta manera para un análisis de sensibilidad de Precios (PSM), el entrevistado emite un concepto acerca de las percepciones que tiene sobre el precio del producto, contestando cada una de las siguientes cuatro preguntas:

$Q_1 =$ ¿A qué precio considera que el producto es barato, le da un buen valor por lo que paga?

$Q_2 =$ ¿A qué precio considera que el producto es caro, pero seguiría comprándolo?

$Q_3 =$ ¿A qué precio considera que el producto es tan costoso que no valdría la pena pagar por él?

$Q_4 =$ ¿A qué precio considera que el producto es tan barato que dudaría de su calidad?

Las revisiones a estas preguntas de precios consisten en hacer unas validaciones lógicas entre las variables para cada uno de los encuestados

- $Q_{1.i} < Q_{2,i}$;
- $Q_{2,i} < Q_{3,i}$;
- $Q_{4,i} < Q_{1,i}$

Donde i representa el individuo i-ésimo, i = 1,2,…n.

El siguiente ejemplo muestra en rojo de la Tabla 11 (Problemas de Coherencia en Precios) donde una de las condiciones no se cumple:

ID	Q1	Q2	Q3	Q4
1	70000	90000	120000	25000
2	55000	70000	90000	20000
3	45000	60000	100000	20000
⋮	⋮	⋮	⋮	⋮
196	40000	60000	100000	40000
197	50000	70000	120000	20000

Tabla 11: Problemas de Coherencia en Precios

En el caso 196 (ID), Q4 es igual al dato de Q1 y no se cumpliría la condición que el encuestado nos dé un precio es tan barato que dudaría de su calidad ya que considera igual al precio barato, le da un buen valor por lo que paga.

Si las validaciones lógicas anteriores no se satisfacen se intenta volver a contactar a la persona para que nuevamente responda por los cuatro precios anteriores. Una práctica común en estos estudios es que si alguna condición lógica no se cumple se elimina el registro lo cual genera un sesgo en las estimaciones que se entregarán de la técnica que en algunos casos no debería ser ignorado.

4. TÉCNICAS DE ANÁLISIS

4.1. Modelo Kano

El método Kano (1984) es un instrumento de análisis muy eficaz al momento de identificar las necesidades o exigencias del cliente, y trasladarlas a ideas sobre los atributos de producto que se consideran importantes para los clientes/usuarios y que permitan entender las características de un nuevo producto.

A finales de 1970, el académico japonés de la universidad de Tokio, Noriaki Kano, extendió el concepto de calidad que se usaba hasta ese entonces, el cual calificaba a la calidad de los productos sobre una sola escala, de "bueno" a "malo"**.** Kano utilizó dos dimensiones para evaluar la calidad**:** el grado de rendimiento de un producto y el grado de satisfacción del cliente que lo utiliza.

El rendimiento de un producto se entiende como la proporción entre el resultado que se obtiene y los medios que se utilizan para lograr el mismo. La satisfacción del cliente que según Philip Kotler, define como "el nivel del estado de ánimo de una persona que resulta de comparar el rendimiento percibido de un producto o servicio con sus expectativas".

Kano (1984) trabaja sobre un plano bidimensional funcionalidad-satisfacción, y definió tres tipos de percepción del cliente de la calidad: obligatoria, unidimensional y atractiva.

Los requerimientos atractivos son aquéllos que, por debajo de cierto umbral de funcionalidad, funcionalidad que se trata de la razón de ser o la esencia del producto, que

tiene características que sin las cuales un producto no sería competitivo y que lo hacen único , mantienen un nivel de satisfacción relativamente bajo y constante, pero que, una vez pasado ese umbral, producen un aumento significativo de la satisfacción. Estos requerimientos también suelen denominarse placenteros (en inglés *delighter*).

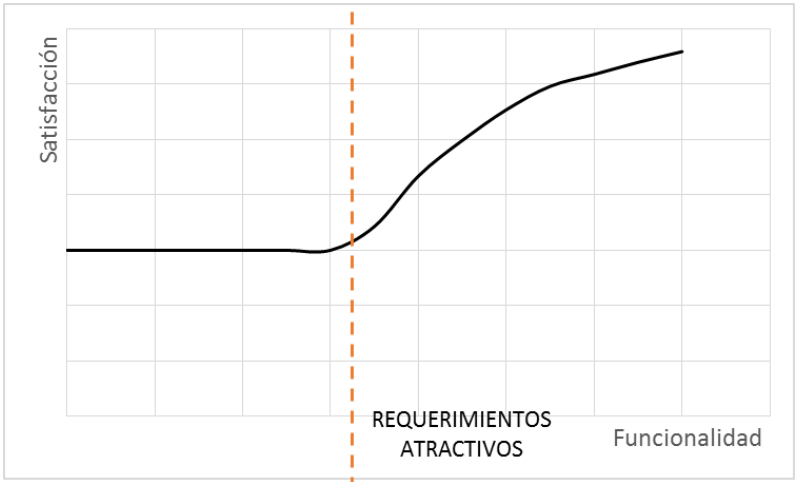


Ilustración 2: Requerimientos Atractivos

Los requerimientos unidimensionales se caracteriza porque la satisfacción aumenta de modo aproximadamente proporcional al nivel de funcionalidad: a mayor funcionalidad, se observa una mayor satisfacción.

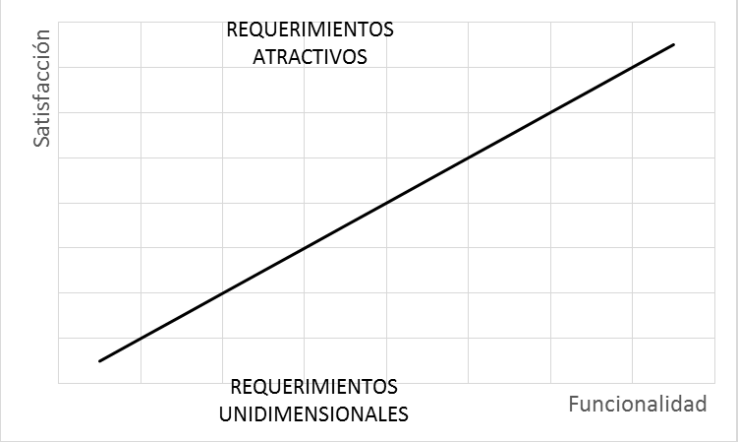


Ilustración 3: Requerimientos Unidimensionales

En los requerimientos se observa que se mantiene la proporcionalidad entre la funcionalidad y la satisfacción del cliente hasta un umbral, pero cuando la funcionalidad sobrepasa cierto límite la satisfacción se mantiene constante.

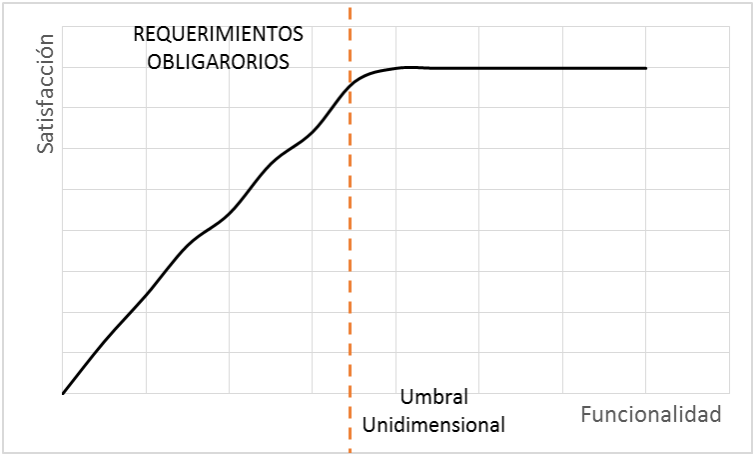


Ilustración 4: Requerimientos Obligatorios

(León Duarte, 2005) amplia el análisis básico de Kano siguiendo al Center for Quality of Management (1993), donde hace un mejor uso de los cuestionarios que Kano (1984) ideó para clasificar las características de un producto y facilitar su diseño.

En la ilustración 5 tomada y adaptada de Center for Quality Manegement (1993), ayuda a comprender el método de Kano, que construye para cada requerimiento del cliente, la relación entre satisfacción y funcionalidad y permite discriminar y clasificar los requerimientos, agrupándolos. En él se han dibujado tres tipos ideales de atributos, en función de la relación entre funcionalidad y satisfacción.

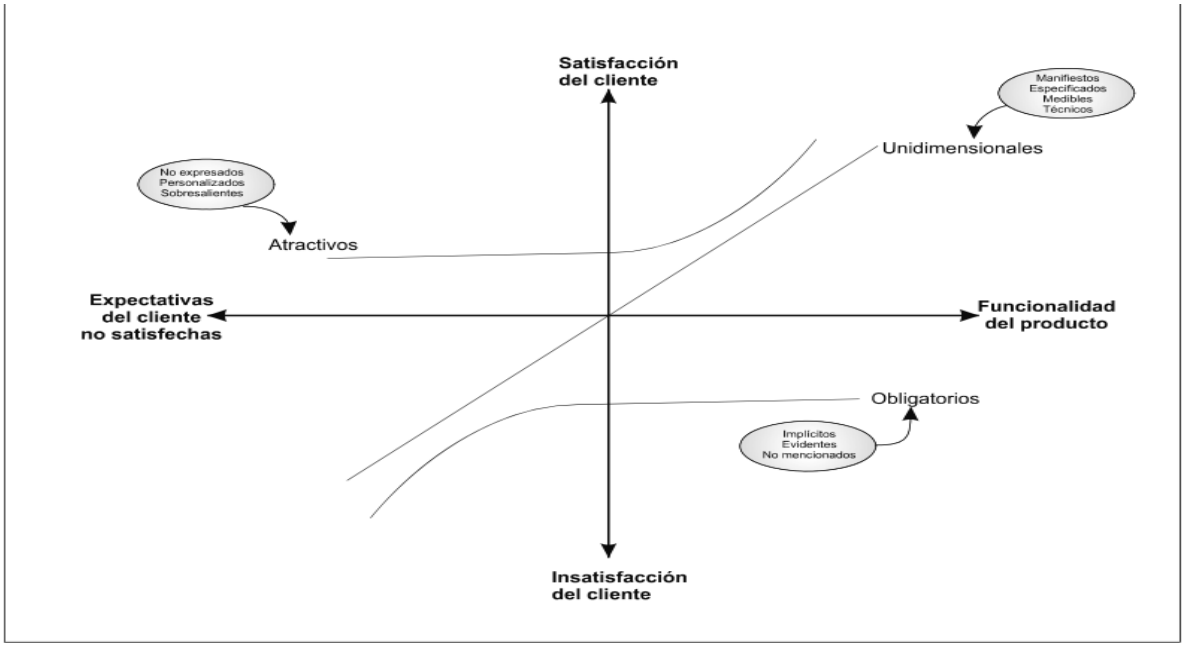


Ilustración 5: Fuente: Center for Quality Manegement (1993).

a. Análisis básico de los datos.

Para aplicar el análisis de Kano en primer lugar se desarrolla una encuesta a clientes y se indagan por diferentes atributos (Ver Palabras Claves) respecto a la satisfacción de un producto.

A continuación se precisa la notación utilizada para el desarrollo del modelo:

$n =$ número de encuestados

$i =$ Individuos participantes en el estudio (corresponde a las filas)

P_k: representa el atributo k – ésimo de la marca. k = 1,2, … A, siendo A el número de atributos a evaluar del producto (marca).

$k_1 =$ representa si el atributo k del producto es funcional (tiene la característica de interés).

k_2: el atributo k del producto es disfuncional (ausencia de la característica de interés).

Kano desarrolló una metodología para clasificar los requerimientos en atractivos, unidimensionales o funcionales teniendo en cuenta la funcionalidad y satisfacción del cliente, para llegar a esta clasificación se plantea dos preguntas para cada atributo k las cuales se miden en escala likert:

- ¿Cómo se siente si la característica k está presente en el producto? (requerimientos funcionales P_{k_1}).

- ¿Cómo se siente si la característica k NO está presente en el producto? (requerimientos disfuncionales P_{k_2}).

Para cada sección, el cliente responde entre 5 opciones, escala Likert. (Ver Palabras Claves). A continuación se ilustra un ejemplo (ver cuadro 1).

REQUERIMIENTOS FUNCIONALES P_{k_1}	Si durante el uso de la base de maquillaje Queda la piel hidratada/humectada, ¿Cómo te sientes?	Me gusta/Me encanta	1
		Es algo básico/Lo mínimo que debe tener	2
		Me da igual/Me da lo mismo	3
		No me gusta, pero lo tolero	4
		No me gusta yNO LO TOLERO	5
REQUERIMIENTOS DISFUNCIONALES P_{k_2}	Si durante el uso de la base de maquillaje No queda la piel hidratada/humectada, ¿Cómo te sientes?	Me gusta/Me encanta	1
		Es algo básico/Lo mínimo que debe tener	2
		Me da igual/Me da lo mismo	3
		No me gusta, pero lo tolero	4
		No me gusta yNO LO TOLERO	5

Cuadro 1 - Preguntas Funcionales y Disfuncionales

Donde $k = 1, 2, \dots A$

Las posibles combinaciones de las respuestas de los requerimientos funcionales y disfuncionales son 25 las cuales se pueden agrupar en las siguientes seis categorías:

I. A: Atractivo
II. O: Obligatorio Básico
III. U: Unidimensional Obligatorio
IV. I: Indiferencia
V. Inv.: Respuesta inversa

VI. D: Respuesta dudosa

En el Cuadro 2 que está basado en la tabla de clasificación de Center for Quality Manegement (1993), se especifican las combinaciones, por ejemplo la categoría atractivo puede obtenerse respondiendo la opción *Me gusta/Me encanta* sobre el requerimiento funcional y las opciones *Es algo básico / lo mínimo que debe tener*, *Me da igual / me da lo mismo*, *No me gusta, pero lo tolero*).

	P_{k_2}					
P_{k_1}		**1**	**2**	**3**	**4**	**5**
	1	**D**	**A**	**A**	**A**	**U**
	2	**Inv.**	**I**	**I**	**I**	**O**
	3	**Inv.**	**I**	**I**	**I**	**O**
	4	**Inv.**	**I**	**I**	**I**	**O**
	5	**Inv.**	**Inv.**	**Inv.**	**Inv.**	**D**

Cuadro 2: Basado en la tabla de clasificación de Center for Quality Manegement (1993).

Una característica se considera *atractiva* si al estar presente en producto al cliente *le gusta / le encanta* mientras que si no está presente la característica no le incomoda, le da igual o cree que es lo mínimo que debería tener, es resumen, son características que sorprenden positivamente si están presenten pero que pueden ser inesperadas.

Una característica se considera obligatoria si su ausencia le genera una gran molestia, pero que no aumenta la satisfacción del cliente si la característica se encuentra en el producto. Como se observa en el cuadro 2 en el cuestionario la característica una característica se clasifica como obligatoria si el cliente responde al requerimiento

disfuncional No me gusta/No lo tolero y además si el cliente no muestra un aumento en su satisfacción al estar presente el atributo pudiendo responder al atributo funcional cualquier opción de las siguientes: Es algo básico / *lo mínimo que debe tener*, *Me da igual / me da lo mismo*, *No me gusta, pero lo tolero*.

Una característica Unidimensional Obligatoria corresponde a clientes que muestran satisfacción bajo la presencia del atributo e insatisfacción bajo la ausencia de este. La categoría atractivo puede obtenerse respondiendo la opción 1 (*Me gusta/Me encanta)* sobre el requerimiento funcional y la opción 5 sobre el requerimiento disfuncional (*No me gusta, y no lo tolero*).

Un cliente puede ser Indiferente a una característica de calidad, esto indica que una mayor o menor funcionalidad respecto a esta característica no se refleja en un aumento o disminución de la satisfacción del cliente y en la Ilustración 1 se representa como una recta paralela al eje horizontal. En este caso el cliente respondería tanto a los requerimientos funcionales como a los disfuncionales con 2, 3 o 4 (respectivamente *Es algo básico / lo mínimo que debe tener*, *Me da igual / me da lo mismo*, *No me gusta, pero lo tolero*).

Una respuesta inversa indica que la interpretación de criterios funcionales y disfuncionales del diseñador es la inversa a la percepción del cliente (lo que la pregunta supone como funcional es percibido como no funcional por quien responde). En este caso sería deseable para el cliente que el producto no tuviera la característica funcional.

En la característica disfuncional el cliente escoge la opción *Me gusta/Me encanta*. En las características funcionales el cliente escoge una de las siguientes opciones *Es algo básico / lo mínimo que debe tener*, *Me da igual / me da lo mismo*, *No me gusta, pero*

lo tolero, No me gusta, y no la tolero, en ninguna de las opciones a excepción de la primera manifiestan satisfacción hacia el atributo. En resumen para el producto sería mejor no tener la característica.

Cuando existe una contradicción en las respuestas a las preguntas, se clasifica en el de respuesta dudosa (ante un par de preguntas complementarias no es razonable contestar “No me gusta, pero lo tolero” a la pregunta funcional y “No me gusta, pero lo tolero” a la disfuncional), de igual forma a la pregunta funcional tampoco es razonable contestar tanto en la pregunta funcional como disfuncional *Me gusta/Me encanta*.

b. Tabla de respuestas.

En la tabla de respuestas se mira la concentración de réplicas propias a cada una de las preguntas del cuestionario, en donde el objetivo es mirar la dispersión de las respuestas.

Autores (Matzler y Hinterhuber, 1998) optan por hacer un gráfico en dos dimensiones con estos datos, en donde también se pueden identificar patrones de agrupaciones de respuestas.

En el cuadro 3 se elaboran índices que ID corresponde al identificador del cliente, P_{k_1}corresponde al valor en escala Likert de la característica funcional del atributo k. P_{k_2}corresponde al valor en escala Likert de la característica disfuncional del atributo k. La columna “Clasificación” corresponde a la síntesis de la información según sea su clasificación: obligatorios básicos (O), unidimensionales obligatorios (U), atractivos (A), los indiferentes (I) y finalmente, los inversos (Inv) y los dudosos (D).

Las columnas siguientes son variables *dummies* las cuales toman el valor de 1 si el cliente clasifica el atributo como obligatoria, unidimensional, atractivo, etc según las reglas previamente explicadas.

Dónde:

$$O_i^k = \begin{cases} 1 & Si\ X = "O" \\ 0 & e.o.c \end{cases}$$

$$U_i^k = \begin{cases} 1 & Si\ X = "U" \\ 0 & e.o.c \end{cases}$$

$$A_i^k = \begin{cases} 1 & Si\ X = "A" \\ 0 & e.o.c \end{cases}$$

$$I_i^k = \begin{cases} 1 & Si\ X = "I" \\ 0 & e.o.c \end{cases}$$

$$Inv_i^k = \begin{cases} 1 & Si\ X = "Inv" \\ 0 & e.o.c \end{cases}$$

$$D_i^k = \begin{cases} 1 & Si\ X = "D" \\ 0 & e.o.c \end{cases}$$

Con k = 1, 2, … A. Siendo A el número de atributos de la marca.

ID	P_{k_1}	P_{k_2}	Clasificación	O_i^k	U_i^k	A_i^k	I_i^k	Inv_i^k	D_i^k
1									
2									
⋮									
i									
⋮									
n									
Sumas				$\sum_{i=1}^{n} O_i$	$\sum_{i=1}^{n} U_i$	$\sum_{i=1}^{n} A_i$	$\sum_{i=1}^{n} I_i$	$\sum_{i=1}^{n} Inv_i$	$\sum_{i=1}^{n} D_i$

Cuadro 3: Sintetizado de la información del análisis básico

En la última columna se ubican las sumas de las anteriores variables *dummies* (Ver Palabras Claves), estas sumas representan el número de encuestados que son

clasificados según los requerimientos que tienen ellos con respecto a la característica del producto.

c. Representación alterna.

Toda la información plasmada en los totales de encuestados clasificados en las seis categorías pueden resumirse en una interpretación alterna de la clasificación de los requerimientos, basada en el incremento de la satisfacción (indicado en la fórmula como *Mejor*) o bien el decremento de la misma (indicado en la fórmula como *Peor*) debida a la colocación o no de una necesidad como característica del producto. Estas fórmulas se obtienen de la idea de ser Mejores que la competencia al satisfacer requerimientos tipo A (Atractivos) y U (Unidimensionales Obligatorios) o bien de la de ser Peores que la competencia al no satisfacer requerimientos tipo U (Unidimensionales Obligatorios) y O (Obligatorios Básicos). En el denominador de ambas fórmulas aparece una sumatoria de las percepciones de atributo A (Atractivos), O (Obligatorios Básicos), U (Unidimensionales Obligatorios) e I (Indiferencia). Se ha suprimido de la sumatoria las percepciones de Inv. (Respuesta inversa) y D (Respuesta dudosa) por su propio carácter impreciso.

$$Mejor = Me_k = \frac{\sum_{i=1}^{n} A_i + \sum_{i=1}^{n} U_i}{\sum_{i=1}^{n} A_i + \sum_{i=1}^{n} U_i + \sum_{i=1}^{n} O_i + \sum_{i=1}^{n} I_i} < 1$$

$$Peor = Pe_k = \frac{\sum_{i=1}^{n} U_i + \sum_{i=1}^{n} O_i}{\sum_{i=1}^{n} A_i + \sum_{i=1}^{n} U_i + \sum_{i=1}^{n} O_i + \sum_{i=1}^{n} I_i} < 1$$

Estos valores se grafican en una escala bidimensional de Mejor y Peor en donde se puede identificar la clasificación del tipo de requerimiento o pregunta (ver gráfico 1).

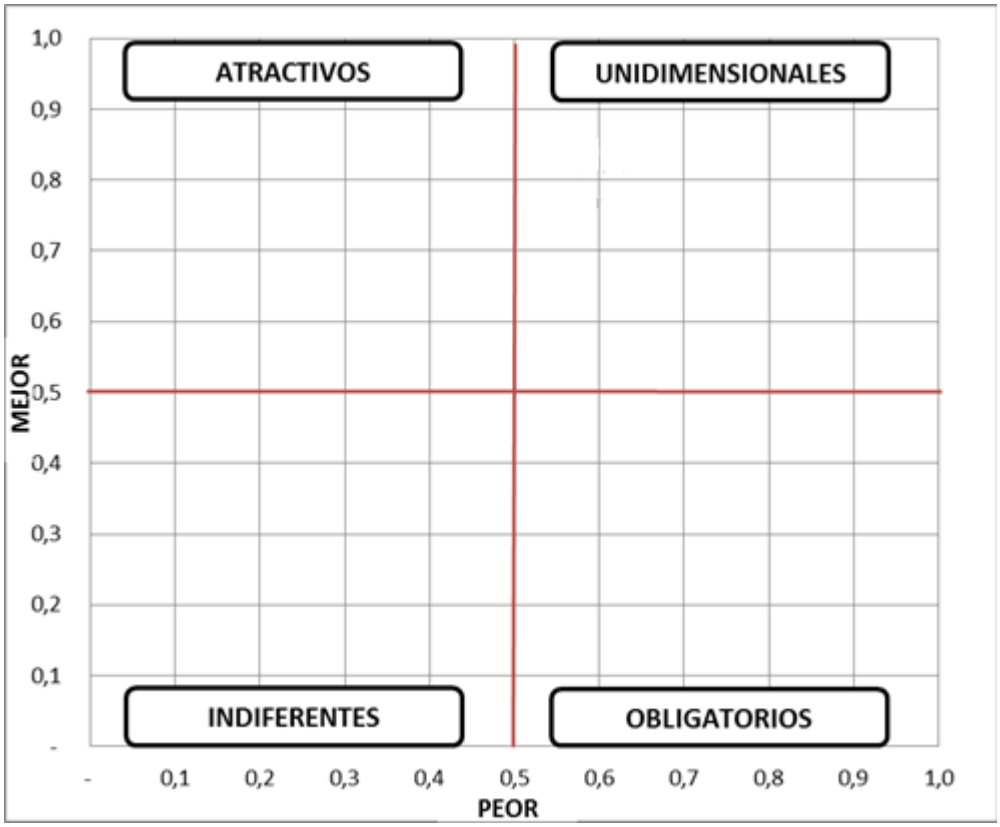


Gráfico 1: Escala bidimensional de Mejor y Peor

En este plano deben graficarse todos los atributos de tal manera que se pueda identificar grupos de atributos que pertenecen a los segmentos, es decir que se debe seguir el mismo procedimiento, para todas los que se tengan.

d. Prueba estadística.

(Fong, 1996) inventó una prueba que se fundamenta en calcular el valor de la diferencia absoluta de las dos frecuencias más votadas de las alternativas (A=Atractivo, O=Obligatorios Básicos, U=Unidimensionales Obligatorios, I=Indiferentes, Inv.=Respuesta inversa, y D=Respuesta Dudosa) y compararlo con el estadístico Q, con el fin de evaluar la significatividad de la clasificación de Kano:

$$Q = \sqrt{\frac{(a+b)(2n-a-b)}{2n}}$$

Donde:

a, b: las frecuencias de las dos observaciones más frecuentes entre A, O, U, I, Inv, o D.

n: número total de respuestas.

De tal forma que:

$$a = \max(\sum_{i=1}^{n} A_i\,, \sum_{i=1}^{n} U_i\,, \sum_{i=1}^{n} O_i\,, \sum_{i=1}^{n} I_i\,, \sum_{i=1}^{n} Inv_i\,, \sum_{i=1}^{n} D_i)_{(6)}$$

$$b = \max(\sum_{i=1}^{n} A_i\,, \sum_{i=1}^{n} U_i\,, \sum_{i=1}^{n} O_i\,, \sum_{i=1}^{n} I_i\,, \sum_{i=1}^{n} Inv_i\,, \sum_{i=1}^{n} D_i)_{(5)}$$

Donde a es el valor máximo de las sumatorias de A, U, O, I, Inv, D, el subíndice (6) indica que es el sexto valor mayor y b en segundo valor máximo de A, U, O, I, Inv, D, con un subíndice (5) que hace referencia que es el quinto mayor valor. El valor Q se

coteja con el de la diferencia absoluta Abs (a-b), y si la diferencia absoluta es menor, muestra que no hay una diferencia significativa entre las dos clasificaciones más frecuentes de cada pregunta, por lo que debe investigarse a fondo, para revelar la presencia de segmentos de mercado reconocibles o problemas en la formulación de la pregunta, en algunos casos debe excluirse el atributo.

Entonces:

$Ho: Q > Abs(a - b)$, No existe diferencia significativa entre las dos clasificaciones más frecuentes.

$Ha: Q \leq Abs(a - b)$, Existe diferencia significativa entre las dos clasificaciones más frecuentes.

e. Cuestionario de atribución de importancia.

Aunque exista concordia en las respuestas, concierta saber si el requerimiento es considerado importante por los clientes. Para ello se utiliza una Parte II de la encuesta, un cuestionario de atribución de importancia, esta calificación es dada en escala de 1 a 9 según el grado de importancia que se asigne a cada requerimiento o pregunta (ver cuadro 4).

Cuestionario de Atribución de Importancia En cada pregunta, marque con un círculo el número de la escala que mejor refleje su opinión									
	Para nada Importante		***Algo Importante***		***Importante***		***Muy Importante***		***Extremo Importante***
Pregunta 1	1	2	3	4	5	6	7	8	9
Pregunta 2	1	2	3	4	5	6	7	8	9
……									
Pregunta X	1	2	3	4	5	6	7	8	9

Cuadro 4: Tomado de Center for Quality Manegement (1993).

Para observar mejor los resultados se elaboró una tabla de unión de los resultados. En esta se concentran los resultados obtenidos para cada uno de los requerimientos en torno a la clasificación conseguida. En base a las fórmulas Mejor y Peor las columnas C1 y C2 se obtienen de multiplicar los valores obtenidos de Mejor y Peor por la Importancia promedio asignada de antemano por el cliente, la cual es a su vez obtenida a partir del promedio de la evaluación de importancia del requerimiento (Si no se realiza la encuesta de Importancia se debe asignar 1 a cada requerimiento o pregunta).

En esta Tabla 12 se observa la concentración de Datos con el peso de importancia.

PREGUNTA	A	O	U	INV	D	I	IMPORTANCIA	MEJOR	PEOR	C1	C2
1	11	6	163	0	0	0	0.55	0.97	0.94	0.53	0.52
2	32	7	126	2	0	13	0.91	0.89	0.75	0.81	0.68

Tabla 12: Ejemplo de tabla de concentración de respuestas. Elaboración Propia. Tomado de Center for Quality Manegement (1993).

4.2. Modelo de Sensibilidad de Precios PSM (Price Sensitivity Meter)

Conocido ampliamente en la literatura anglosajona como *Price Sensitivity Meter (PSM)*. Es una técnica utilizada en la investigación de mercado para determinar en los consumidores preferencias de precios. Fue introducido en 1976 por el economista holandés Peter Van Westendorp. La técnica ha sido utilizada por una amplia variedad de expertos en la investigación de mercado de la industria. El enfoque PSM era una técnica de primera necesidad para hacer frente a problemas de precios durante los últimos 30 años. Históricamente ha sido promovido por muchas asociaciones de investigación de mercados profesionales en su formación y programas de desarrollo profesional tales como la “Sociedad Europea de Opinión e Investigación de Mercados” (*European Society for Opinion and Marketing Research*) (ESOMAR) fundada en 1948.

El enfoque PSM sigue utilizándose ampliamente en la investigación de mercados y los detalles de su implementación se pueden encontrar fácilmente en muchos sitios web relacionados con esta industria.

a. Generalidades

Para aplicar esta técnica se desarrolla un cuestionario a un grupo objetivo que por lo general tiene decisión de compra, dicho cuestionario contiene una serie de preguntas relacionadas con las cualidades del producto y con los precios que el entrevistado asignaría basado en la precepción global que tiene sobre el producto.

El entrevistado emite un concepto acerca de las percepciones que tiene sobre el precio del producto contestando cada una de las siguientes cuatro preguntas:

$Q_1 =$ ¿A qué precio considera que el producto es barato, le da un buen valor por lo que paga?

$Q_2 =$ ¿A qué precio considera que el producto es caro, pero seguiría comprándolo?

$Q_3 =$ ¿A qué precio considera que el producto es tan costoso que no valdría la pena pagar por él?

$Q_4 =$ ¿A qué precio considera que el producto es tan barato que dudaría de su calidad?

b. **Pasos para preparación y el análisis**

Paso 1 validaciones de los precios (ver Problemas coherencia de Precios**)**

Para cada uno de los entrevistados se realizan validaciones lógicas entre las variables:

$Q_{1.i} < Q_{2,i}$; $Q_{2,i} < Q_{3,i}$; $Q_{4,i} < Q_{1,i}$

Donde i representa el individuo i-ésimo, i = 1,2,…n.

Si las validaciones lógicas anteriores no se satisfacen se intenta a contactar al encuestado para que nuevamente responda por los cuatro precios anteriores; si alguna condición lógica no se cumple se debería eliminar el registro completo es decir eliminar las cuatro preguntas, de lo contrario genera un sesgo en las estimaciones que se entregarán de la técnica.

Como se observa en la tabla 2 una vez verificadas las anteriores validaciones para cada uno de los encuestados se lleva a cabo una tabla de frecuencias absolutas con todos los posibles valores de Q_1, Q_2, Q_3 y Q_4. Posteriormente a partir de las anteriores frecuencias se construyen las frecuencias relativas acumuladas.

Paso 2 cálculos de las frecuencias absolutas de los precios

Se generan distribuciones de frecuencia para cada una de las cuatro preguntas mencionadas anteriormente para cada uno de los *P* posibles precios que se presenten en las cuatro preguntas anteriores.

$n_{Q_1,p}$: cantidad de encuestados que piensan que el producto es barato a un precio p que lo compraría, con p =1,2,…P.

$n_{Q_2,p}$: cantidad de encuestados que piensan que el producto es caro a un precio p pero lo compraría, con p =1,2,…P.

$n_{Q_3,p}$: cantidad de encuestados que piensan que el producto es muy caro a un precio a un precio p que no valdría la pena comprarlo, con p =1,2,…P.

$n_{Q_4,p}$ cantidad de encuestados que piensan que el producto es tan barato a un precio p que dudaría de la calidad, con p =1,2,…P.

Cabe resaltar que cada precio que aparece en al menos algunas de las cuatro preguntas es incorporada en la matriz mostrada en la Tabla 13.

Paso 3 cálculos de las frecuencias relativas de los precios

Posteriormente se calcula la tabla de frecuencias relativas:

$f_{Q_1,p}$: proporción o porcentaje de encuestados que piensan que el producto es barato a un precio p que lo compraría, con p =1,2,…P.

$f_{Q_2,p}$: proporción o porcentaje de encuestados que piensan que el producto es caro a un precio p pero lo compraría, con p =1,2,…P.

$f_{Q_3,p}$: proporción o porcentaje de encuestados que piensan que el producto es muy caro a un precio a un precio p que no valdría la pena comprarlo, con p =1,2,…P.

$f_{Q_4,p}$: proporción o porcentaje de encuestados que piensan que el producto es tan barato a un precio p que dudaría de la calidad, con p =1,2,…P.

ID	n_{Q1}	n_{Q2}	n_{Q3}	n_{Q4}	f_{Q1}	f_{Q2}	f_{Q3}	f_{Q4}
1								
2								
⋮								
P								

Tabla 13: Frecuencias relativas y absolutas de precios

Paso 4 cálculos de las frecuencias acumuladas de los precios

En este paso se llevan a cabo las frecuencias absolutas acumuladas (F) para las preguntas Q_2 Q_3. Para estas dos preguntas las frecuencias absolutas se calculan de la siguiente manera:

$$F_{Q_j} = \sum_{p=1}^{P} \frac{f_{Q_j,p}}{n} * 100$$

Siendo p el número de opciones de precio diferentes, con p = 1,2,…P.

Se calculan la diferencias entre uno y la frecuencia relativa acumulada para las preguntas Q_1y Q_4

$$1 - F_{Q_j} = 1 - \sum_{p=1}^{P} \frac{f_{Q_j,p}}{n} * 100$$

Siendo p el número de opciones de precio diferentes, con p = 1,2,…P.

Paso 5 grafica

Se grafica cada una de las curvas obtenidas en un plano precio versus porcentaje de personas.

El método es uno de los más populares para establecer rangos de precios basados en la percepción del comprador final.

Resultados:

OPP: Punto de precio óptimo, es el punto en el cuál el número de respondientes que rechaza el producto por ser demasiado caro $F_{Q_3} * 100$ es igual al número de respondientes que lo rechazan por ser demasiado barato $1 - F_{Q_4} * 100$. Es decir, es la intersección de las curvas $F_{Q_3} * 100\ y\ 1 - F_{Q_4} * 100$ (ver gráfico 3).

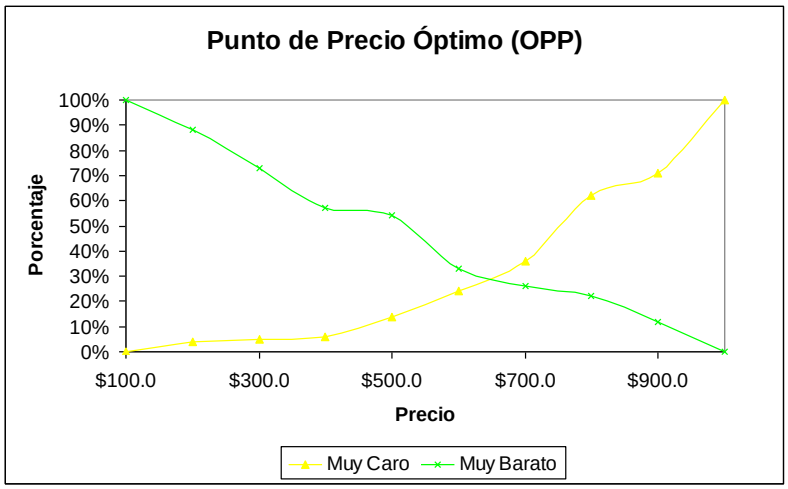


Gráfico 2: Gráfico (OPP)

IPP: Punto de indiferencia, es el punto en el cuál el número de respondientes que consideran el producto barato $1 - F_{Q_1} * 100$ es igual al número de respondientes que consideran el producto caro $F_{Q_2} * 100$. Es decir, es la intersección de las curvas $1 - F_{Q_1} * 100$ y $F_{Q_2} * 100$ (ver gráfico 4).

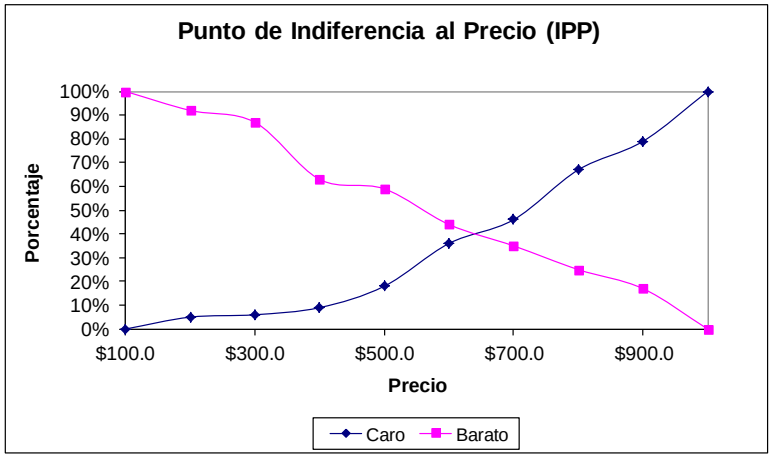


Gráfico 3: Gráfico (IPP)

LMB: Límite marginal barato, punto en el cual el porcentaje de respondientes que encontraron el precio demasiado barato $1 - F_{Q_4} * 100$, es igual al inverso del porcentaje de personas que lo encontraron caro $F_{Q_2} * 100$. Es decir, el porcentaje de personas que consideran que el precio es “no ganga” (ver gráfico 5).

LMC: Límite marginal caro, punto en el cual el porcentaje de respondientes que encontraron el precio demasiado caro $F_{Q_3} * 100$, es igual al inverso del porcentaje de personas que lo encontraron barato $1 - F_{Q_1} * 100$. Es decir, el porcentaje de personas que consideran que el precio es “no caro” (ver gráfico 5).

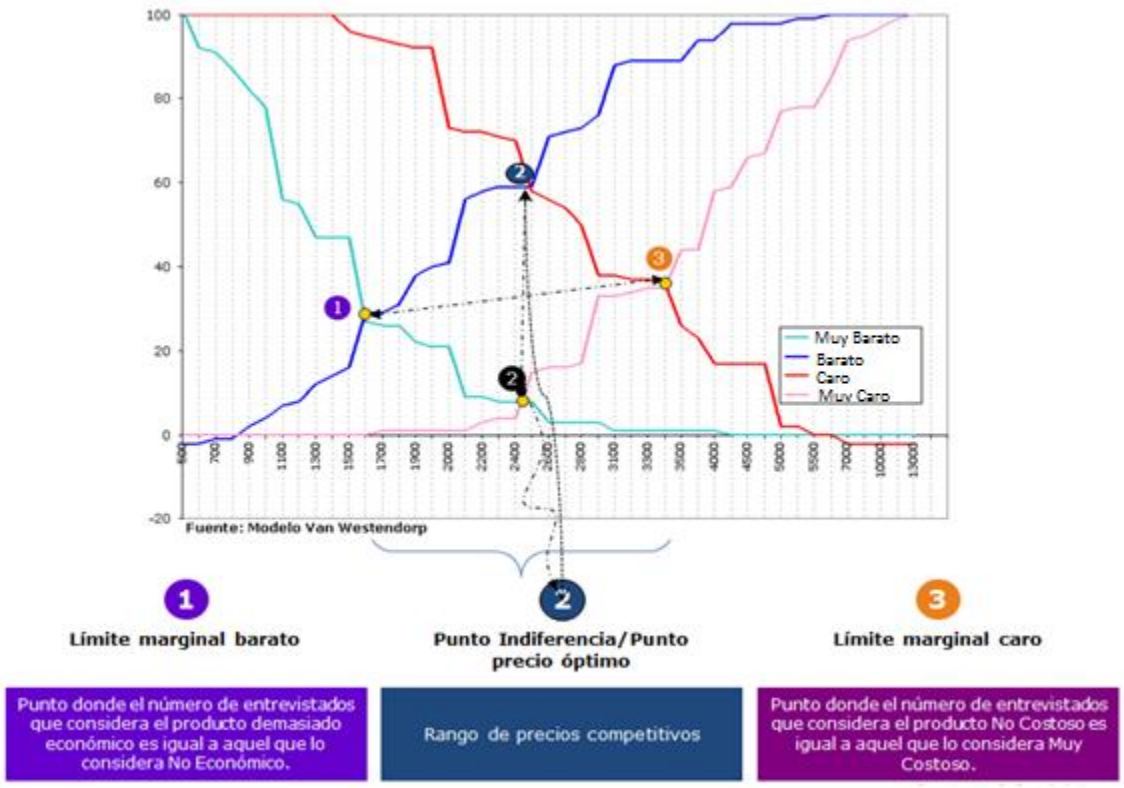


Gráfico 4: Gráfico completo de PSM

El rango entre LMB y LMC conforma el área de precio aceptable para la mayoría de los encuestados.

Como otra manera de mostrar el resultado de PSM para enriquecer el análisis de la información obtenida, se entregan Box-Plot con las respuestas encontradas en cada rango de precios. El diagrama de cajas muestra el percentil 25 y 75 de cada pregunta (Q_1 a Q_4), la mediana, el valor mínimo y máximo, los valores atípicos y la simetría de la distribución (ver gráfico 6)

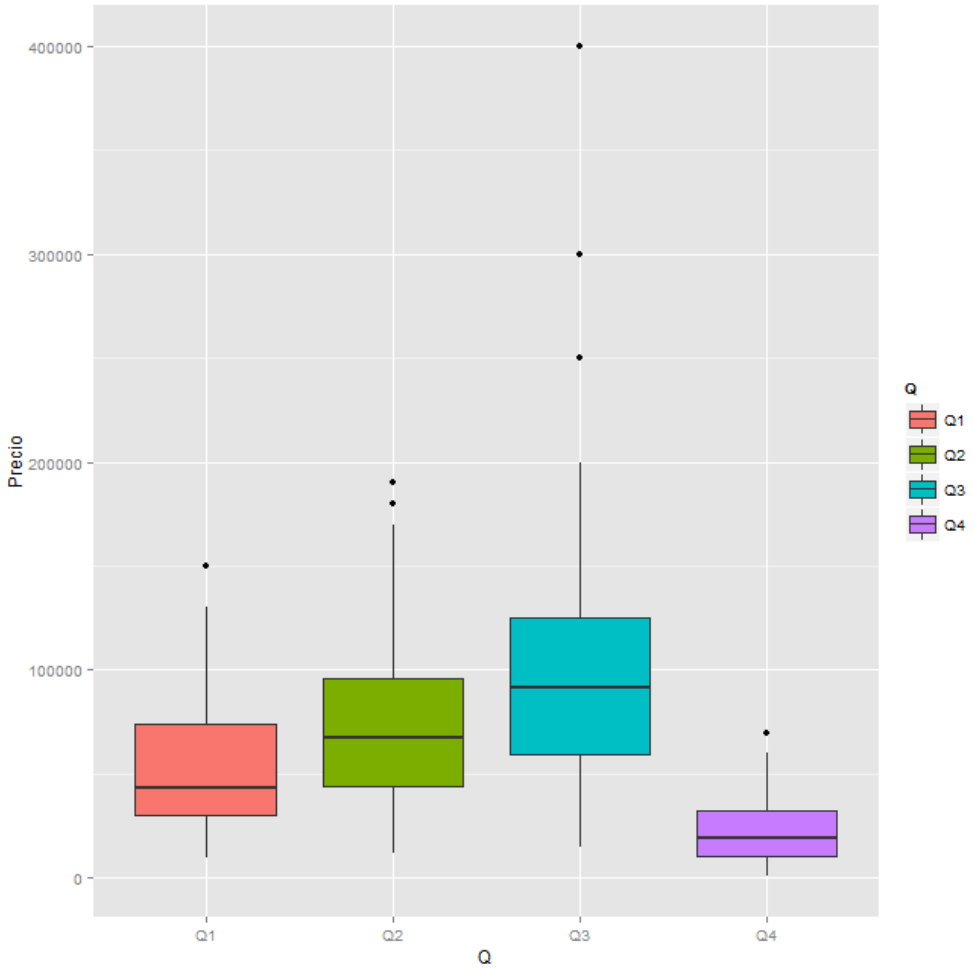


Gráfico 5: Box Plot PSM

4.3. Medidas de similitud en datos Binarios

Los índices de similitud se han venido utilizando en la investigación de mercados para salir al paso con la dificultad hay entre la presencia de un atributo en dos individuos y puede decir sobre el ‘parecido’ de ambos como también la ausencia del atributo.

Para escoger la distancia a utilizar, tendremos en cuenta el tipo de datos con los que estamos trabajando, sin olvidar que en Medidas de Distancia y Similitud trabaja con la matriz de distancias o disimilitudes. Es decir, analizar si se trata de datos cuantitativos, binarios o mixtos.

Un tipo muy común de datos son los datos binarios donde se manifiesta la ausencia o presencia de una serie de características. Los datos binarios suelen ir codificados con ceros y unos. Uno de los más utilizados en mercadeo es:

a. Índice Jaccard

Es un coeficiente de similitud y distancia discreto, este es un estimador cuantitativo que describe el grado de asociación o semejanza entre los elementos comparados, por ejemplo Imagen y el uso, expresado en valor numérico, entre 0 y 1, indicando 1 la similitud máxima.

Muchas veces interesa entender cómo varían juntas dos variables dicotómicas (variables del tipo sí/no o Pertenece/No pertenece).

La correlación es una mala idea en este caso, porque la correlación es una estadística diseñada para datos continuos y para que se comporte de manera “aceptable” es mejor utilizarla con escalas de al menos cinco puntos.

Para investigar la relación entre variables de este tipo utilizamos una estadística que mide la ocurrencia conjunta de las variables. Esta estadística se conoce como el índice de Jaccard (1908).

El índice de Jaccard depende de tres sencillos conteos de incidencia: el número de marcas compartidas entre dos atributos y el número de marcas únicas en cada atributo. Estos conteos son denotados usualmente como *a, b* y *c*, respectivamente (ver tabla 6). El índice Jaccard para el conteo de incidencia entonces es:

$$J_k = \frac{a}{a + b + c}$$

Está basado en la relación entre los atributos y alguna medida de satisfacción, debe ser binarias, Jaccard sólo puede calcularse para dos medidas, que deben ser codificados como 1/0, como sólo se utilizan medidas binarias, no se utilizan medidas escaladas, como los Top Box o Top Two Box (Ver Palabras Claves) se crean a partir de medidas de escala, en estos casos debe recodificarse la escala a binaria para que sea adecuada.

Se debe tener en cuenta que se correlacione los atributos de imagen (medida de categoría no de marcas específicas) con la medida más apropiada.

Se pueden utilizar para dar un indicador a preguntas como:

- Satisfacción Global.
- Intención de Compra.

Los atributos altamente correlacionados con la medida de satisfacción son considerados muy importantes y los atributos con baja correlación o sin correlación son considerados como no importantes.

Las marcas más grandes tendrán mayor impacto en los resultados, proporciona una imagen más realista de la categoría.

Los usos que tiene este índice están en:

- Dar profundidad de aspectos profundos, inconscientes u ocultos de consumidores (insights) en cuanto a la estructura de la categoría.
- Ayudar en los análisis de segmentación.

Que se necesita:

$Z =$ Pregunta asociada a las marcas en donde se pregunta por cada una de ellas, 1 si la tiene 0 si no la tiene (Respuesta Múltiple).

$A =$ Pregunta asociada a los beneficios que quiero comprobar, las respuesta debe ser dada con respecto a Marcas, 1 si la tiene 0 si no la tiene (Respuesta Múltiple).

$k =$ Atributo

$n =$ Tamaño, número total de respuestas de los individuos.

$i =$ Individuo

$m =$ Marcas, m=1,2,…,M

De tal manera que debe haber una relación entre Z y A ya que la posición de las m marcas, debe ser la misma para los dos entonces tenemos que para

$$Z^{(m)} \quad y \quad A_k^{(m)}$$

Donde:

$$A_k^{(m)} = \begin{cases} 1 & si\ la\ marca\ m-ésima\ tiene\ el\ atributo\ k \\ 0 & e.o.c \end{cases}$$

En la Tabla 14 se presenta la estructura de la base de datos:

ID	$Z^{(m)}$	…	$Z^{(M)}$	$A_k^{(m)}$	…	$A_k^{(M)}$
1						
2						
⋮						
i						
⋮						
n						

Tabla 14: Estructura de la Base de Jaccard

Por tal razón:

$$Z^{(m)} = Hace\ referencia\ a\ la\ marca\ m$$

$$Z^{(m)} = \begin{cases} 1 & si\ es\ pertenece\ m \\ 0 & e.o.c \end{cases}$$

Para entender mejor el tipo de preguntas y la estructura que deben usarse para este análisis, puede seguirse el siguiente ejemplo de 100 casos que hacen referencia a un estudio con usuarios de crema dental, donde se indaga por cuatro marcas en específico:

Z = ¿Cuáles de las siguientes marcas podrían cubrir las necesidades de su familia?,

Las marcas por las que se quiere investigar son: 1. JGB, 2. Fluocardent, 3. Fortident, 4. Colgate

A_1= ¿Cuáles de las siguientes marcas asociaría usted estos atributos, el empaque es el más atractivo?, se indaga por las mismas cuatro marcas: 1. JGB, 2. Fluocardent, 3. Fortident, 4. Colgate

A_2= ¿Cuáles de las siguientes marcas asociaría usted estos atributos, el sabor de la crema es agradable?, 1. JGB, 2. Fluocardent, 3. Fortident, 4. Colgate

Las preguntas A_1 y A_2 hacen referencia a dos atributos, para el ejemplo solo se tomaron dos, sin embargo pueden medirse muchos más.

En la base 2 se pude observar la estructura de la información de las preguntas relacionadas en el ejemplo (ver base 1, base completa ANEXOS Excel Base 2):

ID	$Z^{(1)}$	...	$Z^{(4)}$	$A_1^{(1)}$	...	$A_1^{(4)}$	$A_2^{(1)}$	...	$A_2^{(4)}$
1	0	...	1	0	...	0	0	...	0
2	1	...	1	1	...	0	1	...	1
3	0	...	0	1	...	1	1	...	1
4	0	...	0	1	...	1	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	0	...	1	1	...	0	1	...	1
100	0	...	0	0	...	1	1	...	0

Base 1: Base completa ANEXOS Excel Base 2

De tal modo que:

$$Z^{(1)} = Hace\ referencia\ a\ la\ marca\ JGB$$

$$Z^{(1)} = \begin{cases} 1 & si\ es\ "JGB" \\ 0 & e.o.c \end{cases}$$

$$Z^{(2)} = Hace\ referencia\ a\ la\ marca\ Fluocardent$$

$$Z^{(2)} = \begin{cases} 1 & si\ es\ "Fluocardent" \\ 0 & e.o.c \end{cases}$$

$$Z^{(3)} = Hace\ referencia\ a\ la\ marca\ Fortident$$

$$Z^{(3)} = \begin{cases} 1 & si\ es\ "Fortident" \\ 0 & e.o.c \end{cases}$$

$$Z^{(4)} = Hace\ referencia\ a\ la\ marca\ Colgate$$

$$Z^{(4)} = \begin{cases} 1 & si\ es\ "Colgate" \\ 0 & e.o.c \end{cases}$$

La tabla de conteos para cada $Z^{(m)}$ con m=1,2,…,M quedaría de la siguiente manera (ver Tabla 15).

		$A_k^{(m)}$	
		1 (Si)	0 (No)
$Z^{(m)}$	1 (Si)	$a_k^{(m)}$	$b_k^{(m)}$
	0 (No)	$c_k^{(m)}$	–

Tabla 15: Tabla de Conteos

Esto quieres decir que cuando:

$$a_k^{(m)} = \begin{cases} 1 & Z^{(m)} = 1,\ A_k^{(m)} = 1 \\ 0 & e.o.c \end{cases}$$

$$b_k^{(m)} = \begin{cases} 1 & Z^{(m)} = 1,\ A_k^{(m)} = 0 \\ 0 & e.o.c \end{cases}$$

$$c_k^{(m)} = \begin{cases} 1 & Z^{(m)} = 0,\ A_k^{(m)} = 1 \\ 0 & e.o.c \end{cases}$$

Debe contener por lo menos la presencia de una marca para pertenecer al a,b o c; para que pertenezca a a debe contener 1 en Z *y 1* en A para la marca m; de tal manera que para pertenecer a b debe contener en 1 en Z y 0 en A en la marca m y por ultimo para pertenecer a c debe contener 0 en Z y 1 en A para la marca m, en los demás casos es 0 ya que si no pertenece a ninguna marca es que no tienen presencia, esto se debe realizar para cada uno de los entrevistados.

De tal forma que:

$$J_k = \frac{\sum_{m=1}^{M} a_k^{(m)}}{\sum_{m=1}^{M} a_k^{(m)} + \sum_{m=1}^{M} b_k^{(m)} + \sum_{m=1}^{M} c_k^{(m)}}$$

Hace referencia a el índice Jaccard, donde se suman las variables obtenidas en a,b y c de todas las marcas.

Finalmente se calcula el promedio de la variable J_k:

$$\overline{J_k} = \frac{\sum_{i=1}^{n} J_{k,i}}{n}$$

Continuando con el ejemplo se crean las columnas correspondientes a cada marca, en esta base se encuentra todas las marcas para el atributo 1 y los conteos de a, b y c; luego las columnas con las sumas de los conteos, y finalmente se crea el cálculo de la variable J_k (ver base 2, base completa en ANEXOS Excel Base 3):

ID	$Z^{(1)}$	...	$Z^{(4)}$	$A_1^{(1)}$	...	$A_1^{(4)}$	$a_1^{(1)}$	...	$a_1^{(4)}$	$b_1^{(1)}$	...	$b_1^{(4)}$	$c_1^{(1)}$	...	$c_1^{(4)}$	$\sum a_1$	$\sum b_1$	$\sum c_1$	J_1
1	0	...	1	0	...	0	0	...	0	0	...	1	0	...	0	2	1	0	0,7
2	1	...	1	1	...	0	1	...	0	0	...	1	0	...	0	2	1	1	0,5
3	0	...	0	1	...	1	0	...	0	0	...	0	1	...	1	0	1	2	0,0
4	0	...	0	1	...	1	0	...	0	0	...	0	1	...	1	2	0	2	0,5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	0	⋮	0	1	⋮	0	0	⋮	1	⋮	⋮	⋮	⋮
99	0	...	1	1	...	0	0	...	0	0	...	1	1	...	0	0	1	2	0,0
100	0	...	0	0	...	1	0	...	0	0	...	0	0	...	1	0	1	2	0,0

Base 2: Base completa ANEXOS Excel Base 3

Luego se calcula el promedio:

$\overline{J_1} = 0.40$

De la misma manera se realiza para el Atributo 2 (para ejemplo del atributo 2 ver base 3, base completa en ANEXOS Excel Base 4):

Donde el promedio para este atributo es:

$\overline{J_2} = 0.41$

ID	$Z^{(1)}$	...	$Z^{(4)}$	$A_2^{(1)}$	...	$A_2^{(4)}$	$a_2^{(1)}$	...	$a_2^{(4)}$	$b_2^{(1)}$	...	$b_2^{(4)}$	$c_2^{(1)}$	...	$c_2^{(4)}$	$\sum a_2$	$\sum b_2$	$\sum c_2$	J_2
1	0	...	1	0	...	0	0	...	0	0	...	1	0	...	0	1	2	0	0,3
2	1	...	1	1	...	1	1	...	1	0	...	0	0	...	0	2	1	0	0,7
3	0	...	0	1	...	1	0	...	0	0	...	0	1	...	1	0	1	3	0,0
4	0	...	0	0	...	0	0	...	0	0	...	0	0	...	0	2	0	0	1,0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	0	...	1	1	...	1	0	...	1	0	...	0	1	...	0	1	0	1	0,5
100	0	...	0	1	...	0	0	...	0	0	...	0	1	...	0	1	0	2	0,3

Base 3: Base completa ANEXOS Excel Base 4

También podemos darle un ranking según el promedio de cada atributo, donde se ordena del mayor al menor, esto nos daría que entre mayor sea el promedio la asociación es mayor.

Para los dos atributos que se evaluaron en el ejemplo se tiene que (ver tabla 16):

Atributo		Promedio	Ranking
El sabor de la crema es agradable	A_2	0,41	1
El empaque es el más atractivo	A_1	0,40	2

Tabla 16: Ranking Atributos

Donde el primer lugar lo tendrá el que tenga un mayor promedio (A_2 El sabor de la crema es agradable) seguido del siguiente mayor promedio (A_1 El empaque es el más atractivo).

En el caso de tener más atributos se ordenara sucesivamente en el orden jerárquico de mayor a menor, los promedios para los k atributos que se evalúen.

4.4. Análisis de Penalidades (Penalty Analysis)

Es una herramienta que permite recomendar si deben realizar ajustes a las variables organolépticas de un producto y además en qué sentido deben hacerse esos ajustes.

En resumen el análisis de penalidades (*Penalty Analysis*) es un método utilizado en el análisis de datos sensoriales para identificar posibles direcciones en la mejora de los productos, sobre la base de encuestas realizadas a los consumidores o expertos.

Normalmente en las pruebas de producto es necesario generar recomendaciones de producción.

Es muy importante que en producción se pueda identificar si se deben hacer ajustes o no a las fórmulas de los nuevos productos, o saber si los ajustes ya realizados son aceptados por el consumidor.

Un análisis económico en términos de tiempo, costos y esfuerzo, pero además efectivo en términos de separar los resultados sobre un producto de acuerdo a sus características y no como un todo, es conocido como *Penalty Analysis.*

Para llevar a cabo el análisis se requieren los siguientes elementos:

$G =$ Una pregunta de agrado general con los datos de preferencia o puntuaciones de agrado, que corresponden a un índice de satisfacción global para un producto (por ejemplo, la satisfacción global en una escala de 5 puntos para una fragancia o el nivel de dulce de un alimento), pueden presentarse escalas de 7 puntos o 9 puntos también.

Código	Label
1	Es muy desagradable
2	Es desagradable
3	Es ni agradable, ni desagradable
4	Es agradable
5	Es muy agradable

Tabla 17: Ejemplo Pregunta de Agrado de cinco niveles

PJ = Los datos recogidos en variables de *JAR, Just About Right o también llamada Right Factor* (Ver Palabras Claves): Estos corresponden a las calificaciones van de 1 a 5 para una o más características del producto de interés donde el centro contiene el foco, también se pueden presentar en escalas de 7 o 9.

Código	Label
1	Mucho menos dulce del que me gusta
2	Menos dulce del que me gusta
3	Tiene el nivel de dulce que a mi me gusta
4	Más dulce del que me gusta
5	Mucho más dulce del que me gusta

Tabla 18: Ejemplo Pregunta de JAR de cinco niveles

Para las variables de *PJ (JAR)* los datos de la escala se agrupan las categorías 1 y 2, y las categorías 4 y 5, lo que conduce a una escala de tres puntos. Ahora se tiene tres niveles: "No es suficiente", "JAR", y "Demasiado", en el caso de tener 7 o 9 puntos debe recodificarse dejando 3 Niveles, JUSTO en la mitad. Además se utilizara (ver Tabla 19):

$k =$ Atributo

$n =$ Tamaño, número total de respuestas.

$i =$ Individuo

Código	Label	Nuevo Código	
1	Mucho menos Dulce del que me gusta	2	$PJ_2^{(k)}$
2	Menos Dulce del que me gusta		
3	Tiene el nivel justo de Dulce que me gusta	3	$PJ_3^{(k)}$
4	Más Dulce del que me gusta	4	$PJ_4^{(k)}$
5	Mucho más Dulce del que me gusta		

Tabla 19: Recodificación Pregunta JAR de cinco niveles

De tal manera que después de recodificar la base, quedaría con las variables $PJ^{(k)} y\ G$ de la siguiente forma: (ver Tabla 20):

ID	G	$PJ^{(k)}$
1		
2		
⋮		
i		
⋮		
n		

Tabla 20: Estructura de la Base para Penalty Analysis

En el siguiente ejemplo simulado que hace referencia a un estudio de chicles marca X, donde se le dio a probar el nuevo chicle a 100 personas, después de capturada

la información se puede observar cómo se calculan los datos para el análisis (ver base 4, base completa en ANEXOS Excel Base 5):

Las toma la pregunta general (G), en este caso el Agrado y un solo atributo con una escala *JAR* (PJ), de tal forma que:

$G =$ ¿Que tanto le agrada el producto que acaba de consumir?

Contestar según la siguiente escala donde 1 es No le agrada y 5 Le Agrada mucho

$PJ^1 =$ Para usted el nivel de dulce del producto es:

Contestar según la siguiente escala donde 2. Poco, 3. Justo y 4. Mucho, en este caso la escala ya está recodificada acorde al análisis, hay que recordar que de no estar en esta misma escala debe convertirse según Tabla 19 Recodificación Pregunta JAR de cinco niveles.

ID	G	$PJ^{(1)}$
1	2	2
2	5	3
3	1	2
4	3	3
⋮	⋮	⋮
99	4	4
100	1	4

Base 4: Base completa ANEXOS Excel Base 5

Ahora bien se generara una agregación de la variable G dependiendo la respuesta dada en $PJ^{(k)}$, de tal manera que se crean los niveles l que identificara los niveles gusto de la variable $PJ^{(k)}$, de esta manera $PJ_l^{(k)}$ con $l = \{2,3,4\}$ (ver tabla 12).

En general habrá, g niveles equivalentes a las respuestas de la variable G. En la tabla se encuentra considerado el caso en que hay un número igual de observaciones, n_i en cada variable $PJ_l^{(k)}$, sin embargo lo más probable es que no tenga el mismo número de observaciones.

De tal forma que:

$$PJ_l^{(k)} = \begin{cases} g & Si\ PJ^{(k)} = l \\ vacio & e.o.c \end{cases}$$

$$con\ l = \{2,3,4\}$$

ID	G	$PJ_2^{(k)}$	$PJ_3^{(k)}$	$PJ_4^{(k)}$
1				
2				
⋮				
i				
⋮				
n				

Tabla 21: Nueva estructura de la Base de Penalty Analysis

Continuando con el ejemplo del chicle marca X, la nueva estructura quería de la siguiente manera:

ID	G	$PJ_2^{(1)}$	$PJ_3^{(1)}$	$PJ_4^{(1)}$
1	2	2		
2	5		5	
3	1	1		
4	3		3	
⋮	⋮	⋮	⋮	⋮
99	4			4
100	1			1

Base 5: Base completa ANEXOS Excel Base 6

La pregunta PJ^1 se separó en tres variables, $PJ_2^{(1)}$ que se refiere a las personas que calificaron como 2 (Poco) el nivel de dulce del producto, $PJ_3^{(1)}$ que consideran que el nivel de dulce es justo como les gusta y $PJ_4^{(1)}$ que creen que el nivel de dulce es Mucho.

Como lo indica la fórmula:

$$PJ_2^{(1)} = \begin{cases} g & Si\ PJ^{(1)} = 2 \\ vacio & e.o.c \end{cases}; \quad PJ_3^{(1)} = \begin{cases} g & Si\ PJ^{(1)} = 3 \\ vacio & e.o.c \end{cases} \quad y \quad PJ_4^{(1)} = \begin{cases} g & Si\ PJ^{(1)} = 4 \\ vacio & e.o.c \end{cases}$$

Los valores que se colocan en cada variable son los valores correspondientes a la respuesta dada en G (ver ejemplo Base 6, base completa en ANEXOS Excel Base 6).

Ahora bien Siendo n_l número de observaciones de cada nivel l, tenemos que:

$$n = \sum_{l=2}^{4} n_l^{(k)}$$

ID	G	$PJ_2^{(k)}$	$PJ_3^{(k)}$	$PJ_4^{(k)}$
1				
2				
⋮				
i				
⋮				
n		$n_2^{(k)}$	$n_3^{(k)}$	$n_4^{(k)}$

Tabla 22: Nueva estructura de la Base de Penalty Analysis incluyendo n para cada

Nuevamente en el ejemplo de chicles, los valores de $n_l^{(k)}$ pertenecen a los conteos de cada nivel de la variable PJ, es decir la cantidad de datos que contiene cada nivel, en la Base 7 (base completa ANEXOS Excel Base 6) se puede observar los valores pertenecientes a $n_2^{(1)}, n_3^{(1)}\ y\ n_4^{(1)}$:

ID	G	$PJ_2^{(1)}$	$PJ_3^{(1)}$	$PJ_4^{(1)}$
1	2	2		
2	5		5	
3	1	1		
4	3		3	
⋮	⋮	⋮	⋮	⋮
99	4			4
100	1			1
n_l		32	44	24

Base 6: Base completa ANEXOS Excel Base 6

Por lo que:

$$n = \sum_{l=2}^{4} n_l^{(1)}$$

$$n_2^{(1)} = 32, \qquad n_3^{(1)} = 44, \qquad n_4^{(1)} = 24$$

$n =$ 32 + 44 + 24 = 100

Luego se debe obtener la proporción con la siguiente formula:

$$f_l^{(k)} = \frac{n_l^{(k)}}{n}$$

De tal modo que las proporciones de la variable $PJ^{(k)}$ serian:

$$Poco = f_2^{(k)} = \frac{n_2^{(k)}}{n}; \qquad Justo = f_3^{(k)} = \frac{n_3^{(k)}}{n}; \qquad Mucho = f_4^{(k)} = \frac{n_4^{(k)}}{n}$$

Se calculan las medias correspondientes a los niveles l:

$$\overline{PJ}_l^{(k)} = \frac{\sum_{g=1}^{G} PJ_{l,g}^{(k)}}{n_l^{(k)}};\ \text{con}\ g=1,2,\ldots,G$$

$$\overline{PJ}_2^{(k)} = \frac{\sum_{g=1}^{G} PJ_{2,g}^{(k)}}{n_2^{(k)}}; \qquad \overline{PJ}_3^{(k)} = \frac{\sum_{g=1}^{G} PJ_{3,g}^{(k)}}{n_3^{(k)}}; \qquad \overline{PJ}_4^{(k)} = \frac{\sum_{g=1}^{G} PJ_{4,g}^{(k)}}{n_4^{(k)}}$$

En el ejemplo quedaría la tabla (ver Base 7, base completa ANEXOS Excel Base 6):

ID	G	$PJ_2^{(1)}$	$PJ_3^{(1)}$	$PJ_4^{(1)}$
1	2	2		
2	5		5	
3	1	1		
4	3		3	
⋮	⋮	⋮	⋮	⋮
99	4			4
100	1			1
n_l		32	44	24
$f_l^{(1)}$		0,32	0,44	0,24
$\overline{PJ}_l^{(1)}$		2,6	4,3	2,8

Base 7: Base completa ANEXOS Excel Base 6

Donde en las proporciones $PJ^{(1)}$ dependen de cada uno de los niveles de la variable PJ (Nivel de dulce), tenemos que n=100, por lo que las proporciones serían:

$$Poco = f_2^{(1)} = \frac{32}{100} = 0.32; \qquad Justo = f_3^{(1)} = \frac{44}{100} = 0.44; \qquad Mucho = f_4^{(1)} = \frac{24}{100} = 0.24.$$

Y las medias, que corresponden a cada nivel serían:

$$\overline{PJ}_2^{(1)} = \frac{2+1+\cdots+2}{32} = 2.6; \qquad \overline{PJ}_3^{(1)} = \frac{5+3+\cdots+5}{44} = 4.3; \qquad \overline{PJ}_4^{(1)} = \frac{4+\cdots+4+1}{24} = 2.8$$

Después de obtener las proporciones y las medias se calculan las Penalidades para Poco y para Mucho de esta manera:

$$Penalty\ Poco = PP = f_2^{(k)} * \left(\overline{PJ}_2^{(k)} - \overline{PJ}_3^{(k)}\right)$$

$$Penalty\ Mucho = PM = f_4^{(k)} * \left(\overline{PJ}_4^{(k)} - \overline{PJ}_3^{(k)}\right)$$

Reemplazando la formula por los valores obtenidos en el ejemplo de chicles marca X, donde para el PP se multiplica el valor de la proporción correspondiente al nivel 2 (en este caso Poco), por la diferencia entre las medias de los niveles 2 y 3 (Poco y Justo en ese orden), y para PM tenemos que se multiplica el valor de la proporción correspondiente al nivel 4 (en este caso Mucho), por la diferencia entre las medias de 4 y 3 (Mucho y Justo correspondientemente):

$$PP = 0.32 * (2.6 - 4.3) = -0.56$$

$$PM = 0.24 * (2.8 - 4.3) = -0.36$$

En las Penalidades *PP* y *PM* este se identifica con esta guía:

- Menos a (-0.25) = bajo nivel: puede salir de este atributo solo
- Entre (-0,25) y (-0,50) = nivel medio: considere hacer cambios a este atributo
- Mayor a (-0.50) = alto nivel: definitivamente realizar cambios en este atributo

En la realización de análisis pena (*Penalty Analysis*) puede experimentar un número positivo. Los números positivos se ven como algo positivo y no debe ser

cambiado. Por ejemplo, en un batido calificaciones de los consumidores puede resultar en una puntuación de penalización positiva para espesor, que es un positivo debido a que el consumidor querer un producto más espeso.

Teniendo en cuenta los valores obtenidos en el ejemplo de chicles, los ajustes que se deben hacerse corresponden a estos dos valores -0.56 en *PP* y -0.36 en *PM*; ¿pero cómo saber si la el ajuste que se debe realizar es significativo?

Para esto se realiza una prueba múltiple de HSD Tukey en la cual se pondrá a prueba las medias de Justo con Poco y Justo con Mucho.

Tukey (1953) propuso un procedimiento para testar la hipótesis nula, el máximo de α siendo exactamente el nivel global de significancia,

El test de HSD Tukey utiliza la distribución de la estadística de amplitud en la forma de Student:

$$q = \frac{\bar{J}^k_{max} - \bar{J}^k_{min}}{\sqrt{\frac{MQ_E}{n}}}$$

Siendo: $\bar{J}^k_{max} - \bar{J}^k_{min}$, la mayor y menor media respectivamente.

$$T_\alpha = \frac{q_\alpha(\alpha, f)}{\sqrt{2}} \sqrt{MQ_E \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Donde MQ_E es el error cuadrático medio y f el número de grados de libertad, asociado con MQ_E y $q_\alpha(\alpha, f)$ se halla en las tablas estadísticas de Tukey.

Si el valor absoluto de la diferencia entre dos medias fuera mayor de T_α entonces H_0 debe ser rechazado.

$H_0 = \overline{PJ}_l^{(k)} = \overline{PJ}_{l+1}^{(k)}$

$H_1 = \overline{PJ}_l^{(k)} \neq \overline{PJ}_{l+1}^{(k)}$; Para todo $l \neq l+1$

En el ejemplo de chicles el HSD Tukey tiene la siguiente prueba la hipótesis:

1. $H_0 = \overline{PJ}_2^1 = \overline{PJ}_3^1$ ó $2.6 = 4.3$

$H_1 = \overline{PJ}_2^1 \neq \overline{PJ}_3^1$ ó $2.6 \neq 4.3$

Y

2. $H_0 = \overline{PJ}_4^1 = \overline{PJ}_3^1$ ó $2.8 = 4.3$

$H_1 = \overline{PJ}_4^1 \neq \overline{PJ}_3^1$ ó $2.8 \neq 4.3$

Para los dos casos existen evidencias estadísticas para rechazar H_0 es decir que existen diferencias significativas entre las medias.

Gráficamente se vería de esta forma:

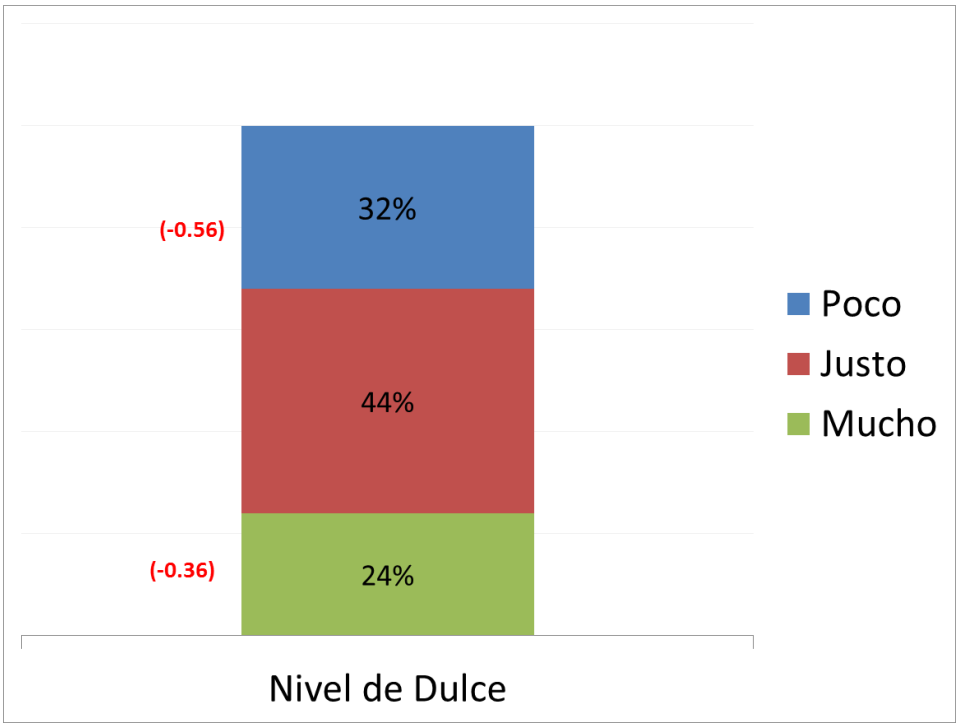


Gráfico 6: Gráfico de Barras PSM

Las penalizaciones en rojo entre paréntesis y las proporciones a las que correspondan en cada valor.

También puede presentarse un mapa de dispersión con los dos valores de las penalidades en el eje y la proporción correspondiente en el eje X:

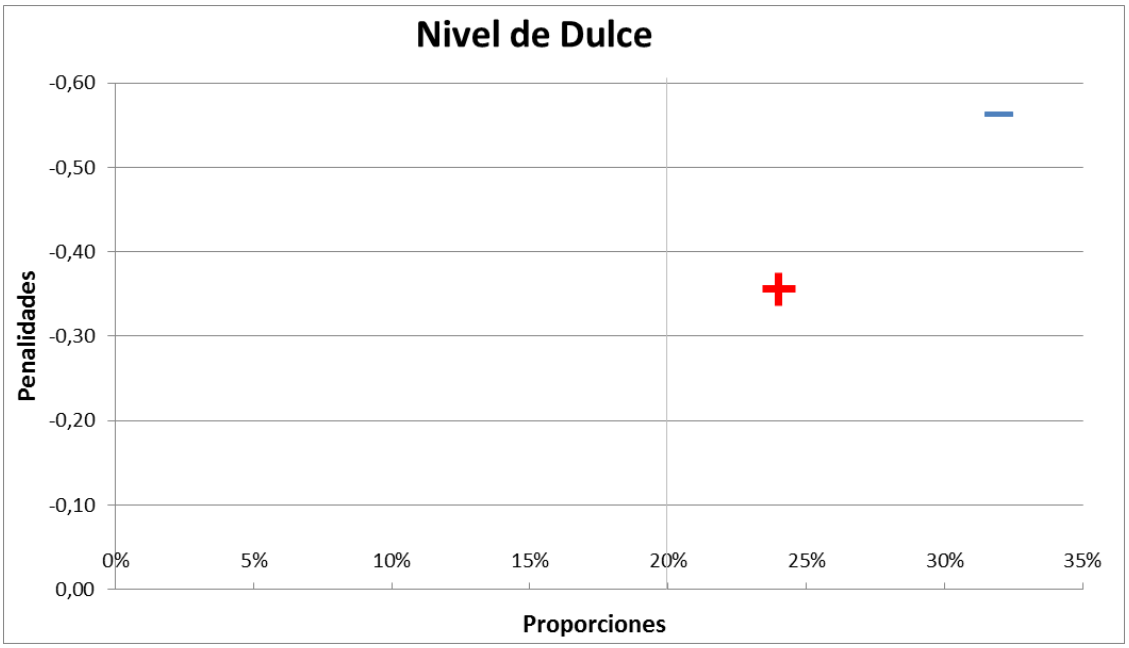


Gráfico 7: Gráfico Penalidades

En el plano deben ir cada uno de los atributos a evaluar para su mejor análisis.

Una consideración que se debe tener en cuenta es que análisis debería ejecutarse para los atributos donde al menos 20% o más de los encuestados califiquen en ya sea en $PJ_2^{(k)}$ o $PJ_4^{(k)}$.

DESARROLLOS COMPUTACIONALES

1. USO DE LAS FUNCIONES PROBLEMAS BASE DE DATOS

1.1. Uso Problemas reestructuración de bases con rotación de productos (problem_rotations)

Función para la reestructuración de bases de datos con dos alternativas, los argumentos son:

```
R> problem_rotations (data, pos_id, first_alt, start_first_alt, end_first_alt, second_alt,
+  start_second_alt, end_second_alt,  start_gen_quest, end_gen_quest)
```

Donde:

data	dataframe que contiene las variables de interés
pos_id	posición del identificador, generalmente ID
first_alt	columna que contiene el espacio en orden de la primera alternativa a evaluar
start_first_alt	columna donde comienza las respuestas que pertenecen a la primera alternativa
end_first_alt	columna donde termina las respuestas que pertenecen a la primera alternativa
second_alt	columna que contiene el espacio en orden de la segunda alternativa a evaluar
start_second_alt	columna donde comienza las respuestas que pertenecen a la segunda alternativa
end_second_alt	columna donde termina las respuestas que pertenecen a la segunda alternativa
start_gen_quest	columna donde comienza las respuestas que entran completas sin reestructurar, preguntas generales

end_gen_quest	columna donde termina las respuestas que entran completas sin reestructurar, preguntas generales

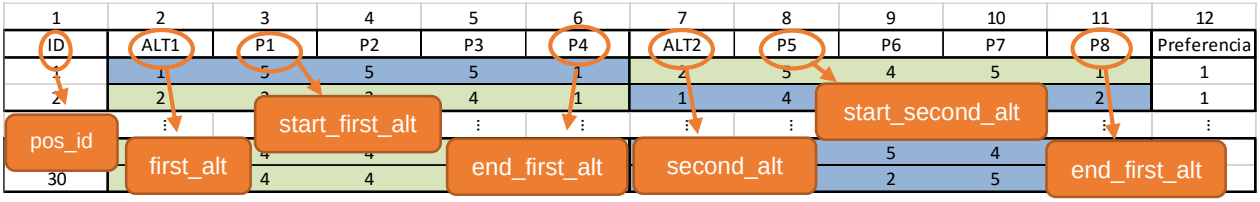
Tabla 23: Problemas reestructuración de bases con rotación de productos

Un ejemplo con una base simulada con 30 casos que sirve para ilustrar mejor el proceso que sigue la función:

ID	ALT1	P1	P2	P3	P4	ALT2	P5	P6	P7	P8	Preferencia
1	1	5	5	5	1	2	5	4	5	1	1
2	2	2	3	4	1	1	4	4	4	2	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	2	4	4	4	2	1	4	5	4	1	1
30	2	4	4	4	1	1	4	2	5	1	1

Cuadro 5: Ejemplo Problemas reestructuración de bases con rotación de productos

La base está compuesta de dos bloques, que van desde ALT1 hasta Preferencia, en color azul el bloque correspondiente al producto 1, y el bloque verde hace referencia al producto 2, de tal forma que se pueden identificar los argumento para la función:



Cuadro 6: Identificación de Argumentos función (problem_rotations)

Para el ejemplo el dataframe utilizado “Problemas Bases Problemas Reestructurar.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Problemas Bases Problemas Reestructurar.csv",
+ header=TRUE, sep=";",dec=",")
```

Identificados los argumentos de la función, en la parte superior aparece la posición de cada variable, para luego continuar con el reemplazo en la función:

```
R> problem_rotations(data = Base, pos_id = 1, first_alt = 2, start_first_alt = 3,
+ end_first_alt = 6, second_alt = 7, start_second_alt = 8, end_second_alt = 11,
+ start_gen_quest = 12, end_gen_quest = 12)
```

La función busca unificar la base de tal manera que los dos bloques queden en uno solo, de la siguiente manera:

ID	ALT1	P1	P2	P3	P4	Preferencia
1	1	5	5	5	1	1
1	2	5	4	5	1	1
2	1	4	4	4	2	1
2	2	2	3	4	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	1	4	5	4	1	1
29	2	4	4	4	2	1
30	1	4	2	5	1	1
30	2	4	4	4	1	1

Cuadro 7: Nueva estructura Problemas reestructuración de bases con rotación de productos

La salida de la función es un dataframe reestructurado según las especificaciones que se le asignaron:

Output:

```
   ID ALT1 P1 P2 P3 P4 Preferencia
1   1    1  5  5  5  1          1
2   1    2  5  4  5  1          1
```

3	2	1	4	4	4	2		1
4	2	2	2	3	4	1		1
…								
57	29	1	4	5	4	1		1
58	29	2	4	4	4	2		1
59	30	1	4	2	5	1		1
60	30	2	4	4	4	1		1

1.2. Uso Problemas Respuestas Múltiples (identify_rm_dup_km)

El uso de la función sirve para validación en preguntas múltiples para evitar duplicados en estas, los argumentos son:

```
R> identify_rm_dup_km (data, prefix_know, sep_know, identify_dup = T,
+ rm_dup = F, missing_codes = NA)
```

Donde:

prefix_know	prefijo de identificación de las variable antes de sep_know
sep_know	separador de las respuestas múltiples, identifica en donde se separa cada respuesta múltiple
identify_dup	vector de salida identificando las variables duplicadas - TRUE o FALSE, predeterminada TRUE
rm_dup	remover duplicados - FALSE o TRUE, predeterminada FALSE
missing_codes	tratamiento de los datos perdidos tales como no sabe, no responde, predeterminado NA
data	dataframe que contiene las variables de interés

Tabla 24: Problemas Respuestas Múltiples

El programa valida que no existan duplicados dentro de una pregunta con respuestas múltiples, hay que tener en cuenta que en missing_codes se asignan los códigos que hacen referencia a valores perdidos o códigos como 99 (No sabe, No responde) el cual podría generar un duplicado sin que en realidad lo sea, en el caso de tener más de una código de estos debe asignarse de la siguiente manera c(98, 99).

Un ejemplo con 30 datos simulados, que muestran una pregunta múltiple que va desde la P1.1 hasta la P1.6, la pregunta por ejemplo puede ser las marcas de gaseosas que conoce:

ID	P1.1	P1.2	P1.3	P1.4	P1.5	P1.6
1	21	26	30	21		
2	12					
3	12					
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	9					
29	12					
30	6	7	12	30		

Cuadro 8: Ejemplo Problemas Respuestas Múltiples

Para identificar los argumentos en la base se puede apreciar en la siguiente tabla:

ID	P1.1	P1.2	P1.3	P1.4	P1.5	P1.6
1	21	26	30	21		
2	12					
3	12					
⋮				⋮	⋮	⋮
28	9					
29	12					
30	6	7	12	30		

prefix_know

sep_know

Cuadro 9: Identificación de Argumentos función (identify_rm_dup_km)

De esta manera el prefix_know es “P1” y el sep_know es “.”, la base se encuentra en un dataframe llamado “Problemas Bases Problema Respuestas Multiples.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Problemas Bases Problema Respuestas Multiples.csv",
+ header=TRUE, sep=";",dec=",")
```

Ahora bien, identificadas las variables para el uso de la función se tiene que:

```
R> identify_rm_dup_km(data = Base, prefix_know = "P1",
+ sep_know = ".", identify_dup=TRUE, rm_dup = TRUE, missing_codes = NA)
```

Para el ejemplo se utilizó los identify_dup=TRUE, que genera un vector de salida ($is.duplic_know) identificando las variables duplicadas, de utilizar FALSE no existirá tal salida, el mismo caso pasa con rm_dup = TRUE, genera un dataframe quitando los duplicados, es decir solo deja un dato en caso de estar duplicado.

La salida es una lista que contiene $is.duplic_know (esta salida está condicionada), $dup_varnames y $know_matrix_nodup (esta salida está condicionada).

```
Output: $is.duplic_know
 [1]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[11] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[21] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

La salida indica que en el primer caso encontró un duplicado, si se observa en la base efectivamente existe un duplicado en la fila 1:

ID	P1.1	P1.2	P1.3	P1.4	P1.5	P1.6
1	21	26	30	21		
2	12					
3	12					
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	9					
29	12					
30	6	7	12	30		

Cuadro 10: Ejemplo Problemas Respuestas Múltiples

Output: $dup_varnames

[1] "'P1.1RM','P1.4RM'" "" "" "" "" ""

[7] "" "" "" "" "" ""

[13] "" "" "" "" "" ""

[19] "" "" "" "" "" ""

[25] "" "" "" "" "" ""

Esta salida identifica en donde se encuentran los duplicados, en la P1.1 y en la P1.4.

Output: $know_matrix_nodup

	P1.1	P1.2	P1.3	P1.4	P1.5	P1.6
1	21	26	30	NA	NA	NA
2	12	NA	NA	NA	NA	NA
3	12	NA	NA	NA	NA	NA

…

28	9	NA	NA	NA	NA	NA
29	12	NA	NA	NA	NA	NA
30	6	7	12	30	NA	NA

Dataframe de salida que elimina el duplicado encontrado en la fila 1 en la variable P1.6.

1.3. Uso Problema de datos faltantes (identify_missing)

El uso de esta función valida cuando existen datos faltantes comparando entre dos variables generalmente múltiples, los argumentos son:

R> identify_missing (data, prefix_know, sep_P_know, Index_know, sep_I_know)

Donde:

data	dataframe que contiene las variables de interés
prefix_know	prefijo de identificación de las variable antes de sep_P_know
sep_P_know	separador de las respuestas múltiples para el prefix_know
Index_know	prefijo de identificación de las variable Índice antes de sep_I_know
sep_I_know	separador de las respuestas múltiples para el Index_know

Tabla 25: Problema de datos faltantes

Generalmente se realiza con preguntas de respuesta múltiple, pero también se puede utilizar con respuestas únicas, se realiza por medio de comparación de dos matrices de tal manera que los dos grupos de variables prefix_know e Index_know.

Un ejemplo de 30 casos simulados, que muestran un estudio de investigación de mercados, donde se investigó sobre cuatro marcas (pregunta múltiple), se asignó colores para identificar la marca y las siguientes dos preguntas P1 y P2 (preguntas múltiples), hacen referencia a preguntas derivadas de las marcas que contestaron en las variables MARCA.

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5	5	4	5	2		
2	1	2	3		4	4	2		4	2	2	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	1	2	5		4	4	4		4	1	2	
30	1	2	3		3	2	3		4	3	3	

Cuadro 11: Ejemplo Problema de datos faltantes

Los argumentos de la función indican que deben existir dos matrices que comparar:

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5				2		
2	1	2	3		4	4				2	2	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	1	2	5		4	4	4		4	1	2	
30	1	2	3		3	2	3		4	3	3	

Index_know

prefix_know

sep_P_know

Cuadro 12: Identificación de Argumentos función (identify_missing)

Se identificaron los parámetros de la función, MARCA será el Índice y P1 el Prefijo, sin embargo sep_I_know no se encuentra, por lo que debe asignarse en la función “”, que indica que no existe un separador en la pregunta Índice.

Para el ejemplo el dataframe utilizado “Base Kano.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Problemas Bases Problema Falta de Información.csv",
+ header=TRUE, sep="; ",dec=",")
```

Ahora bien se utiliza la función para las variables de MARCA y P1:

```
R>  identify_missing(data = Base, prefix_know = "P1", sep_P_know = ".",
+ Index_know = "MARCA", sep_I_know = "")
```

La salida de la función muestra una lista que contiene $logical_missing, $data_missing y $position_missing:

```
Output: $logical_missing
     MARCA1 MARCA2 MARCA3 MARCA4
 [1,]   TRUE   TRUE   TRUE  FALSE
 [2,]   TRUE   TRUE   TRUE   TRUE
 …
[29,]   TRUE   TRUE   TRUE   TRUE
[30,]   TRUE   TRUE   TRUE   TRUE
```

La salida 'logical_missing' identifica en la matriz de Index_know los valores que están faltantes y que no deberían serlo, se asigna un FALSE en la matrix para identificarlos, como se puede identificar en la salida en la fila 1 existe un faltante en la MARCA4, en la base se puede también apreciar en color rojo el faltante:

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4
1	1	2	5		5	5	5	4
2	1	2	3		4	4	2	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	1	2	5		4	4	4	
30	1	2	3		3	2	3	

Cuadro 13: Ejemplo Problema de datos faltantes

Output: $data_missing

ID MARCA1 MARCA2 MARCA3 MARCA4 P1.1 P1.2 P1.3 P1.4 P2.1 P2.2 P2.3 P2.4 Val

1 1 1 2 5 NA 5 5 5 4 5 2 NA NA TRUE

2 2 1 2 3 NA 4 4 2 NA 4 2 2 NA FALSE

La salida 'data_missing' es un dataframe que tiene cargada en data la cual contiene una variable al final "Val" que indica el caso donde presenta el dato faltante, contiene un TRUE en la fila que presenta algún faltante cuando no debería tenerlo.

Output: $Position_missing

Row Column

1 1 MARCA4 , P1.4

La salida 'Position_missing' data.frame que indica en la columna "Row" la fila donde se encuentra el dato faltante y en "Column" el nombre de las columnas que presenta el dato.

1.4. Uso Problemas con duplicidad de información (iden_dup)

El uso de la función Identifica duplicidad de la información entre filas, que sirve para detectar encuestas duplicadas, los argumentos de la función son:

```
R> iden_dup (data, vars, type_ident = "complete")
```

Donde:

data	dataframe que contiene las variables de interés
vars	variables expuestas a identificación de duplicados, se debe marcar comienzo y fin
type_ident	tipo de duplicado que muestra, "fromStart" no muestra el primer duplicado, "fromLast" no muestras el ultimo duplicado, "complete" muestra todos duplicados

Tabla 26: Problemas con duplicidad de información

El programa identifica los duplicados que se encuentran entre las filas, que generalmente son encuestas, con vars se debe marcar comienzo y fin, en posiciones 1:33 o identificando las variables por los nombres de ellas, por ejemplo c("P12","P13").

Cuando se identifica type_ident "fromStart" no muestra el primer caso duplicado, "fromLast" no muestra el último caso duplicado, "complete" muestras todos, si no se marca ninguna por defecto saca "complete".

Un ejemplo con 30 casos simulados, que está compuesta por un ID (Identificador de encuesta), tres variables múltiples MARCA, P1 y P2, donde se muestra la siguiente base:

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5	5		5	2	5	
2	1	2	3		4	4	2		4	2	2	
3	1	2	3	5	4	5	4	4	4	2	4	3
4	2	3	5		2	3	4		4	4	4	
5	1	2	5		2	2	2		4	4	3	
1	1	2	5		5	5	5		5	2	5	
6	1	3			4	4			4	2		
7	2	3	5		2	2	2		4	4	3	
8	1	2	5		3	2	3		4	5	4	
2	3	4			4	5			5	1		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	1	2	5		4	4	4		4	1	2	
30	1	2	3		3	2	3		4	3	3	

Cuadro 14: Problemas con duplicidad de información

La base anterior de ejemplo esta almacenada en un dataframe llamado “Problemas Bases Problema Duplicados.csv”, ahora quedara recopilado en un objeto llamado Base:

```
R> Base <- read.table("Problemas Bases Problema Duplicados.csv",
+ header=TRUE, sep=";",dec=",")
```

Las variables que se utilizaran para identificar si existen o no duplicados son de la columna 1 a la 13.

1	2	3	4	5	6	7	8	9	10	11	12	13
ID	**MARCA1**	**MARCA2**	**MARCA3**	**MARCA4**	**P1.1**	**P1.2**	**P1.3**	**P1.4**	**P2.1**	**P2.2**	**P2.3**	**P2.4**
1	1	2	5		5	5	5		5	2	5	
2	1	2	3		4	4	2		4	2	2	
3	1	2	3	5	4	5	4	4	4	2	4	3

Cuadro 15: Identificación de Argumentos función (iden_dup)

Una vez identificadas las variables para el uso de la función se tiene que:

```
R> iden_dup(data = Base, vars = 1:13, type_ident = "complete")
```

La salida es una lista que muestra $ident_duplicate y $vars_duplicate.

```
Output: $ident_duplicate
[1]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
[11] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[21] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

La salida 'ident_duplicate' es un vector que identifica con TRUE el caso donde encuentre un duplicado, como se puede apreciar en la base, efectivamente se encuentran duplicadas las dos encuestas, la de la fila 1 y la fila 10:

ID	MARCA1	MARCA2	MARCA3	MARCA4	P1.1	P1.2	P1.3	P1.4	P2.1	P2.2	P2.3	P2.4
1	1	2	5		5	5	5		5	2	5	
2	1	2	3		4	4	2		4	2	2	
3	1	2	3	5	4	5	4	4	4	2	4	3
4	2	3	5		2	3	4		4	4	4	
5	1	2	5		2	2	2		4	4	3	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	2	5		5	5	5		5	2	5	

Cuadro 16: Ejemplo Problemas con duplicidad de información

```
Output: $vars_duplicate
ID MARCA1 MARCA2 MARCA3 MARCA4 P1.1 P1.2 P1.3 P1.4 P2.1 P2.2 P2.3 P2.4
1   1     1     2     5     NA    5    5    5    4    5    2    NA   NA
10  1     1     2     5     NA    5    5    5    4    5    2    NA   NA
```

La salida ‘vars_duplicate’ muestra en la data cargada solo las variables que se encuentran duplicadas, dependiendo del type_ident que se escoja, ya que se escogió “complete”, muestra los dos duplicados, al elegir “fromStart” mostrara solamente la encuesta 10, y si por el contrario se escoge "fromLast", se mostrara solamente la encuesta 1.

1.5. Uso Problemas de Coherencia (compareLists)

El uso de la función que sirve para comparar datos entre dos matrices de datos, sus argumentos son:

```
R> compareLists (data, start_mult1, last_mult1, start_mult2, last_mult2,
+ compare = "2in1", incomparables = NA)
```

Donde:

data	dataframe que contiene las variables de interés
start_mult1	columna donde comienza la primera matriz de datos
last_mult1	columna donde termina la primera matriz de datos
start_mult2	columna donde comienza la segunda matriz de datos
last_mult2	columna donde termina la segunda matriz de datos
compare	permite comparar las matrices, ya sea de “2in1” que es predefinido y compara de la matriz 2 a la matriz 1, si por el contario es “1in2” compara la matriz 1 en la matriz 2
incomparables	tratamiento de los datos perdidos tales como no sabe, no responde, ninguno, que no deben ser comparados, predeterminado NA

Tabla 27: Problemas de Coherencia

La salida de esta función muestra las inconsistencias, comparando los números que están en las matrices de datos.

Un ejemplo de 30 datos simulados, que van desde la encuesta 1000 hasta la 1035, la base de datos contiene un ID (Identificador de la encuesta), dos pregunta múltiple P1 y P2, la P1 hace referencia la cantidad de marcas que recuerda de productos de belleza,

la P2 hace referencia a la cantidad de marcas que ha usado en los últimos 3 meses de las marcas que ha mencionado en P1.

ID	P1.RU	P1.OMR1	P1.OMR2	P1.OMR3	…	P1.OMR12	P2R1	P2R2	P2R3	P2R4	…	P2R12
1000	6	1	2				2	3	6	714		
1002	6	1	2	3			1	3	6			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1034	1	2	3				1	2	3			
1035	3	1	6				2	3				

Cuadro 17: Ejemplo Problemas de Coherencia

La bese están identificadas con color cada una de las marcas, esto con el fin de distinguir cuales de las marcas de P2 no están en P1 como lo indica la instrucción, ahora bien se deben, identificar la posición de las variables que pertenecen a las matrices que se incluirán en la función:

1	2	3	4	5		14	15	16	17	18		26
ID	P1.RU	P1.OMR1	P1.OMR2	P1.OMR3	…	P1.OMR12	P2R1	P2R2	P2R3	P2R4	…	P2R12
1000	6	1	2				2	3	6	714		
1002	6	1	2	3			1	3	6			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1034	1	2	3				1	2	3			
1035	3	1	6				2	3				

start_mult1 · last_mult1 · start_mult2 · last_mult2

Cuadro 18: Identificación de Argumentos función (compareLists)

Para el ejemplo el dataframe utilizado “Problemas Bases Problema Coherencia.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Problemas Bases Problema Coherencia.csv",
+ header=TRUE, sep=";",dec=",")
```

Ahora bien, identificadas las variables para el uso de la función se tiene que:

```
R> compareLists(data = Base, start_mult1 = 2, last_mult1 = 14,
+ start_mult2 = 15, last_mult2 = 26, compare = "2in1" , incomparables = c(99,97))
```

La salida de corrresponde a una lista donde se puede observar $is.consistent y $problems.

Los resultados del ejemplo son los siguientes:

```
Output: $is.consistent
 [1] 1 0 0 1 1 1 1 1 1 1 1 0 0 1 0 0 1 0 0 0 0 1 0 1 0 0 1 1 0 1
```

La salida ‘is.consistent’ es un vector que indica con 1 y 0 donde se encuentra la inconsistencia, 1 cuando tiene inconsistencia, 0 cuando no, efectivamente existe varios problemas entre los que se pueden apreciar en color rojo la celda de la base de ejemplo:

ID	P1.RU	P1.OMR1	P1.OMR2	P1.OMR3	…	P1.OMR12	P2R1	P2R2	P2R3	P2R4	…	P2R12
1000	6	1	2				2	3	6	714		
1002	6	1	2	3			1	3	6			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1034	1	2	3				1	2	3			
1035	3	1	6				2	3				

Cuadro 19: Ejemplo Problemas de Coherencia

En la P2R2 y en P2R4 de la encuesta 1000, se encuentra las marcas 3 y 714 las cuales no fueron mencionadas en P1, lo mismo para en la encuesta 1035, la marca 2 indicada en P2R1 no fue señalada en P1.

Output: $problems

[1] "","P2R2","","P2R4"'

[2] "","","""

….

[29] "","","""

[30] "'P2R1',""

La salida ‘problems’ es un vector que contiene el nombre de la variable que contiene la inconsistencia, como efectivamente se podría ver en la base indica que las variables P2R2 y P2R4 tiene un problema, también la P2R1 en la fila 30 que pertenece a la encuesta 1035.

1.6. Uso Problemas de Coherencia en Precios (Val_Prices)

El uso de esta función está encaminado a la Validación de Precios que sirve para aplicar posteriormente el PSM (*Prices Sensitivity Model*), los argumentos de la función son:

```
R> Val_Prices(key, QC, QE, QTE, QTC, data)
```

Donde:

key	Indicador, generalmente es el ID de la encuesta, o puede ser cualquier otra lleve
QC	pregunta de precio considera que el producto es barato, le da un buen valor por lo que paga
QE	pregunta de precio considera que el producto es caro, pero seguiría comprándolo
QTE	pregunta precio considera que el producto es tan costoso que no valdría la pena pagar por él
QTC	pregunta precio considera que el producto es tan barato que dudaría de su calidad
data	dataframe que contiene las variables de interés

Tabla 28: Problemas de Coherencia en Precios

La función realiza una validación de la coherencia de precios, donde se verifica que QC sea menor QE, que QE sea menor que QTE y que QTC sea menor que QC.

Un ejemplo con datos de un nuevo desodorante masculino que va a salir al mercado, con 400 datos simulados de personas, las cuales contestaron las siguientes cuatro preguntas con respecto al precio del producto que probaron:

Q_1= ¿A qué precio considera que el producto que probó es barato, le da un buen valor por lo que paga?

Q_2= ¿A qué precio considera que el producto que probó es caro, pero seguiría comprándolo?

Q_3= ¿A qué precio considera que el producto que probó es tan costoso que no valdría la pena pagar por él?

Q_4= ¿A qué precio considera que el producto que probó es tan barato que dudaría de su calidad?

ID	Q1	Q2	Q3	Q4
1	70000	90000	120000	25000
2	55000	70000	90000	20000
3	45000	60000	100000	20000
⋮	⋮	⋮	⋮	⋮
196	40000	60000	100000	40000
⋮	⋮	⋮	⋮	⋮
400	60000	85000	100000	50000

Cuadro 20: Ejemplo Problemas de Coherencia en Precios

Dadas las preguntas del cuestionario pueden identificar, en la base como se utiliza la función:

key= ID

QC = Barato = Q_1

QE = Caro = Q_2

QTE = Muy Caro = Q_3

QTC = Muy Barato = Q_4

Primero el dataframe utilizado “Base PSM.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Base PSM.csv", header=TRUE, sep=";", dec=",")
```

Seguido se reemplaza las preguntas de la base en la función Val_Prices:

```
R> Base <- Val_Prices(key = ID, QC = Q1, QE = Q2, QTE = Q3, QTC = Q4, data = Base)
```

La salida es un dataframe que muestra las encuestas que presentan algún problema en la validación.

Output:

```
  ID   QC    QE    QTE   QTC QC<QE QE<QTE QTC<QC Validation
196 196 40000 60000 100000 40000  TRUE   TRUE  FALSE      FALSE
```

Esta salida contiene el *ID* que identifica que encuesta presenta problema, además de las columnas *QC, QE, QTE* y *QTC*, están la *QC<QE* columna cuyo resultado arrojara un FALSE cuando la cumpla la condición que QC sea menor *QE, QE<QTE* columna cuyo resultado arrojara un FALSE cuando la cumpla la condición que *QE* sea menor *QTE*, también *QTC<QC* columna cuyo resultado arrojara un FALSE cuando la cumpla la condición que *QTC* sea menor *QC* y por ultimo *Validation* columna de la base que identifica si por lo menos existe un dato que presente un problema con la validación de Precios, marca FALSE si por lo menos en alguna columna *"QC<QE", "QE<QTE"* , *"QTC<QC"* presenta un FLASE, los resultados del ejemplo muestran que efectivamente existe un problema entre los precios de QTC=Q4 y QC=Q1, es decir que no se cumple

que QTC sea menor que QC, también se puede apreciar en la base, que las celdas de color rojo tiene el mismo precio:

ID	Q1	Q2	Q3	Q4
1	70000	90000	120000	25000
2	55000	70000	90000	20000
3	45000	60000	100000	20000
⋮	⋮	⋮	⋮	⋮
196	40000	60000	100000	40000
⋮	⋮	⋮	⋮	⋮
400	60000	85000	100000	50000

Cuadro 21: Ejemplo Problemas de Coherencia en Precios

2. USO DE TÉCNICAS DE ANÁLISIS

2.1. Uso Modelo Kano (KA)

La función en R distingue los siguientes argumentos:

R> KA(QF, QD, data)

Donde:

QF	atributo funcional, tenencia de la característica de interés en KA
QD	atributo disfuncional (ausencia de la característica de interés en KA
data	dataframe que contiene las variables de interés

Tabla 29: Modelo Kano (KA)

Para apreciar mejor el uso de esta técnica de análisis se utilizara un ejemplo de un estudio de un producto de base de maquillaje con datos simulados:

ID	P_{1_1}	P_{1_2}
1	1	4
2	1	4
3	1	4
4	1	5
⋮	⋮	⋮
29	1	5
30	1	4

Cuadro 22: Ejemplo uso Modelo Kano (KA)

En total son 30 casos, la base está compuesta de las variables ID, P_{1_1}, P_{1_2}; ahora bien para identificar las variables funcionales y disfuncionales, se puede mirar las preguntas que se realizaron en el cuestionario:

	A continuación encontrará parejas de preguntas , las cuales responden a un mismo tema, para cada una de las parejas de preguntas usted deberá marcar con una X la opción (Entre: Me Gusta/Me Encanta a No me gusta /No lo tolero). **(E. MOSTRAR TARJETA KANO)**					
		Me gusta/Me encanta	Es algo básico/Lo mínimo que debe tener	Me da igual/Me da lo mismo	No me gusta, pero lo tolero	No me gusta y NO LO TOLERO
P1,1	Si durante el uso de la base de maquillaje **Queda la piel hidratada/humectada**, ¿Cómo te sientes?	1	2	3	4	5
P1,2	Si durante el uso de la base de maquillaje **No queda la piel hidratada/humectada**, ¿Cómo te sientes?	1	2	3	4	5

Ilustración 6: Estructura preguntas uso Modelo Kano (KA)

El atributo de interés hace referencia a la Hidratación/Humectación de la piel, puede identificar la pregunta funcional es:

QF = Si durante el uso de la base de maquillaje Queda la piel hidratada/humectada, ¿Cómo te sientes?.

Y la pregunta disfuncional:

QD = Si durante el uso de la base de maquillaje No queda la piel hidratada/humectada, ¿Cómo te sientes?.

Para el ejemplo el dataframe utilizado “Base Kano.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Base Kano.csv", header=TRUE, sep=";",dec=",")
```

Ahora bien, identificadas las variables para el uso de la función se tiene que:

```
R> KA(QF = "P1_1", QD = "P1_2", data = Base)
```

La salida de corrresponde a una lista donde se puede observar $points y $test_difference_sig.

Los resultados del ejemplo son los siguientes:

```
Output: $points
   Better     Worse
0.8333333 0.6000000
```

Los valores obtenidos en la salida $points, hacen relación con el dato de mejor (Better) y peor (Worse) valores que son puntos de coordenadas que se pueden graficar en un plano e indican en el cuadrante que se ubica el atributo, en este caso Hidratación/humectación de la piel.

```
Output: $test_difference_sig
Q <= abs(a-b)
      TRUE
```

El test que mide la diferencia en este analisis es estre el valor absoluto (a-b) y Q, donde el resultado obtenido “TRUE” indica que existe diferencia significativa entre las dos frecuencias más ocurrentes del ejercicio, por lo que el resultado debería presentarse.

El KA solo aplica para dos variables una funcional y otra disfuncional es decir un solo atributo, mientras que el KA_multiple lo realiza para mas de un atributo los cuales deben estar en la base de datos en forma ordenada por parejas QF1, QD1, QF2, QD2 sucesivamente.

Para ver mejor el manejo del KA_multiple se tomara una ejemplo con una base de con 180 casos simulados, de una fragancia nueva que va a salir al mercado, la base contiene 37 atributos ordenados en pareja funcionales y disfuncionales, los atributos estan definidos según las cualidades del producto, por ejemplo:

- El aroma del producto es agradable
- El envace viene con dosificador; entre otros más atributos.

ID	P1_1	P1_2	P2_1	P2_2	…	P37_1	P37_2
1	1	5	1	4	…	1	4
2	1	5	1	4	…	1	2
3	1	5	1	4	…	1	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
179	1	5	1	5	…	1	4
180	1	5	1	5	…	1	5

Cuadro 23: Ejemplo uso Modelo Kano (KA)

Los argumentos para esta función son:

```
R> KA_multiple(prefix, sep , data)
```

Donde:

Prefix	Prefijo identificador de las variables que se deben analizar en KA_multiple
sep	identifica donde se separa una variable de otra en KA_multiple, el valor predefinido es "_"
data	dataframe que contiene las variables de interés

Tabla 30: uso Modelo Kano (KA_multiple)

El "sep" el valor predefinido para separar las preguntas antes de prefijo es "_", pero también puede ser ".", "-", o un símbolo que en la data identifique el cambio variable.

Identificados los argumentos se indica la indica la función de KA_multiple:

R> KA_multiple(prefix = "P", sep = "_" , data = Base)

Para apreciar de donde viene los argumentos de la base de datos se tiene que:

ID	P1_1	P1_2	P2_1	P2_2	…	P37_1	P37_2
1	1	5	1	4	…	1	4
2	1	5	1	4	…	1	2
3	1	5	1	4	…	1	4
⋮	prefix	sep	⋮	⋮	⋮	⋮	⋮
179	1	5	1	5	…	1	4
180	1	5	1	5	…	1	5

Cuadro 24: Identificación de Argumentos función (KA_multiple)

La salida es la misma que con KA, solo que esta vez se muestra todos los atributos:

Output: $points

```
P1  0.9666667 0.9388889
P2  0.8876404 0.7471910
P3  0.8722222 0.5944444
      …

P35 0.8770950 0.4916201
P36 0.8926554 0.3841808
P37 0.8547486 0.3910615
```

En este caso están los 37 atributos que se asignaron en el análisis

Output: $test_difference_sig

```
Q <= abs(a-b)
P1            TRUE
P2            TRUE
P3            TRUE
      …
```

```
P35             TRUE
P36             TRUE
P37             TRUE
```

También el test de diferencias indica con “TRUE” y “FALSE” para cada uno de los atributos cuando existe o no diferencia entre las calificaciones más altas de cada uno de los atributos.

Para complementar el análisis se grafican los $points de los atributos en un plano:

```
R> salida <- KA_multiple(prefix = "P", sep = "_", data = Base)
R> plot(salida$points[,2], salida$points[,1], pch = 20, col ="gray60", xlab = "Worse",
   +  ylab = "Better", xlim=c(0, 1), ylim=c(0, 1), main = "Kano Analysis")
R> abline(h = 0.5, v = 0.5, col = "red")
R> text(salida$points[,2]-0.03, salida$points[,1],labels = rownames(salida$points),
   + cex = 0.70)
```

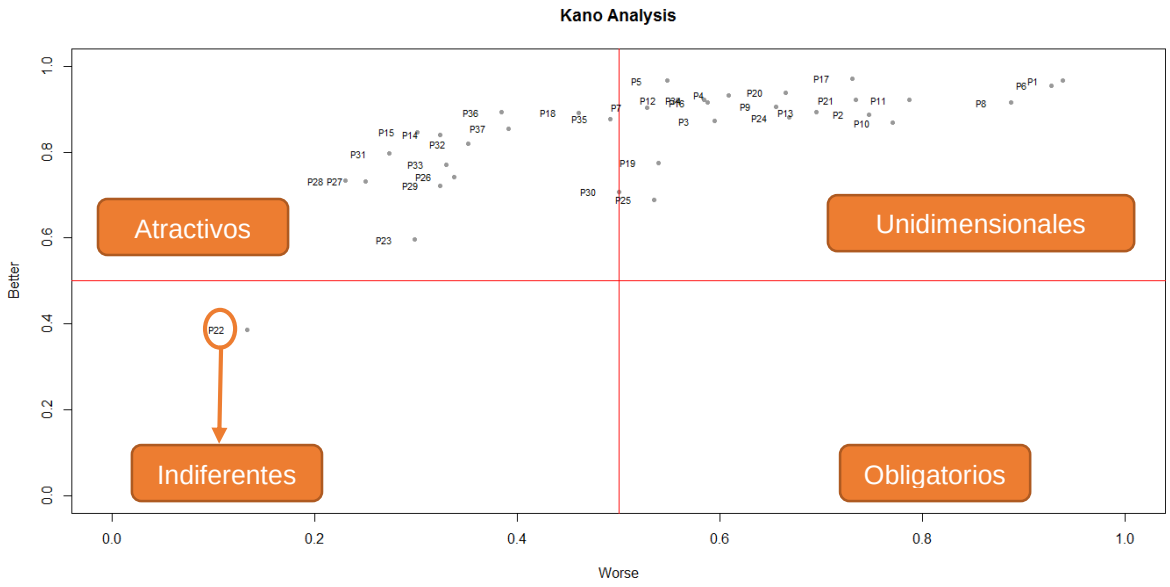


Gráfico 8: Ejemplo uso Modelo Kano (KA_multiple)

En el mapa están ubicados cada uno de los atributos evaluados en el cuadrante al que corresponde, por ejemplo el atributo que corresponde al P22 se encuentra en el cuadrante I (Indiferentes).

2.2. Uso Modelo de Sensibilidad de Precios PSM (PSA)

El uso de la función de Prices Sensitivity Model (*PSM*), se utiliza para para determinar en los consumidores preferencias de precios, los argumentos son los siguientes:

R> PSA(QC, QE, QTE, QTC, data)

Donde:

QC	pregunta de precio considera que el producto es barato, le da un buen valor por lo que paga
QE	pregunta de precio considera que el producto es caro, pero seguiría comprándolo
QTE	pregunta precio considera que el producto es tan costoso que no valdría la pena pagar por él
QTC	pregunta precio considera que el producto es tan barato que dudaría de su calidad
data	dataframe que contiene las variables de interés

Tabla 31: Uso Modelo de Sensibilidad de Precios PSM (PSA)

Un ejemplo con datos de un tratamiento facial que va a salir al mercado, después de probar el producto:

ID	Q1	Q2	Q3	Q4
1	70000	90000	120000	25000
2	55000	70000	90000	20000
3	45000	60000	100000	20000
4	60000	90000	140000	15000
⋮	⋮	⋮	⋮	⋮
399	50000	80000	120000	35000
400	60000	85000	100000	50000

Tabla 32: Ejemplo Uso Modelo de Sensibilidad de Precios PSM (PSA)

Con 400 datos simulados de personas, las cuales contestaron las siguientes cuatro preguntas con respecto al precio del producto que probaron:

$Q_1 =$ ¿A qué precio considera que el producto de tratamiento facial que probó es barato, le da un buen valor por lo que paga?

$Q_2 =$ ¿A qué precio considera que el producto de tratamiento facial que probó es caro, pero seguiría comprándolo?

$Q_3 =$ ¿A qué precio considera que el producto de tratamiento facial que probó es tan costoso que no valdría la pena pagar por él?

$Q_4 =$ ¿A qué precio considera que el producto de tratamiento facial que probó es tan barato que dudaría de su calidad?

Dadas las preguntas del cuestionario pueden identificar, en la base como se utiliza la función:

QC = Barato = Q_1

QE = Caro = Q_2

QTE = Muy Caro = Q_3

QTC = Muy Barato = Q_4

De tal manera que se puede usar la función PSA, primero el dataframe utilizado “Base PSM.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Base PSM.csv", header=TRUE, sep=";", dec=",")
```

Seguido se reemplaza las preguntas de la base en la función PSA:

```
R> Base <- PSA(QC = Q1, QE = Q2, QTE = Q3, QTC = Q4, data = Base)
```

La salida de la función PSA es una tabla que contiene:

Output: $`Table PSM`

	Q	nC	nE	nTE	nTC	fC	fE	fTE	fTC	FC	FE	FTE	FTC
1	1000	0	0	0	1	0.0000	0.0000	0.0000	0.0025	100.00	0.00	0.00	99.75
2	3000	0	0	0	1	0.0000	0.0000	0.0000	0.0025	100.00	0.00	0.00	99.50
	…												
57	300000	0	0	1	0	0.0000	0.0000	0.0025	0.0000	0.00	100.00	99.75	0.00
58	400000	0	0	1	0	0.0000	0.0000	0.0025	0.0000	0.00	100.00	100.00	0.00

La variable Q en la tabla indica el precio, y las demás variables expresan las frecuencias relativas y absolutas que deben graficarse para identificar los puntos OPP: Punto de precio óptimo, IPP: Punto de indiferencia, LMB: Límite marginal barato y LMC: Límite marginal caro:

```
R>  tr <- PSA(QC = Q1, QE = Q2, QTE = Q3, QTC = Q4, data = Base)
R>  options(scipen = 111)
R>  plot(tr$`Table PSM`$Q, tr$`Table PSM`$FC, type="l",lty=1, col="firebrick1",
+     xlim=range(min(tr$`Table PSM`$Q),max(tr$`Table PSM`$Q)),
+     ylim=c(0, 100),main="Price Sensitivity Analysis Plot", xlab="Price",
+    ylab="Porcent")
R> lines(tr$`Table PSM`$Q, tr$`Table PSM`$FE, type="l",pch=1, lty=1,
 +  col="dodgerblue")
```

```
R> lines(tr$`Table PSM`$Q, tr$`Table PSM`$FTE, type="l",pch=1, lty=1,
+  col="lightseagreen")
R> lines(tr$`Table PSM`$Q, tr$`Table PSM`$FTC, type="l",pch=1, lty=1,
+ col="tan1")
R> legend("topright", inset=.05, title="Price", c("Cheap","Expensive",
+ "Too expensive","Too cheap"), lty=c(1,1,1,1),
+   col=c("firebrick1", "dodgerblue", "lightseagreen", "tan1"), cex = 0.5)
```

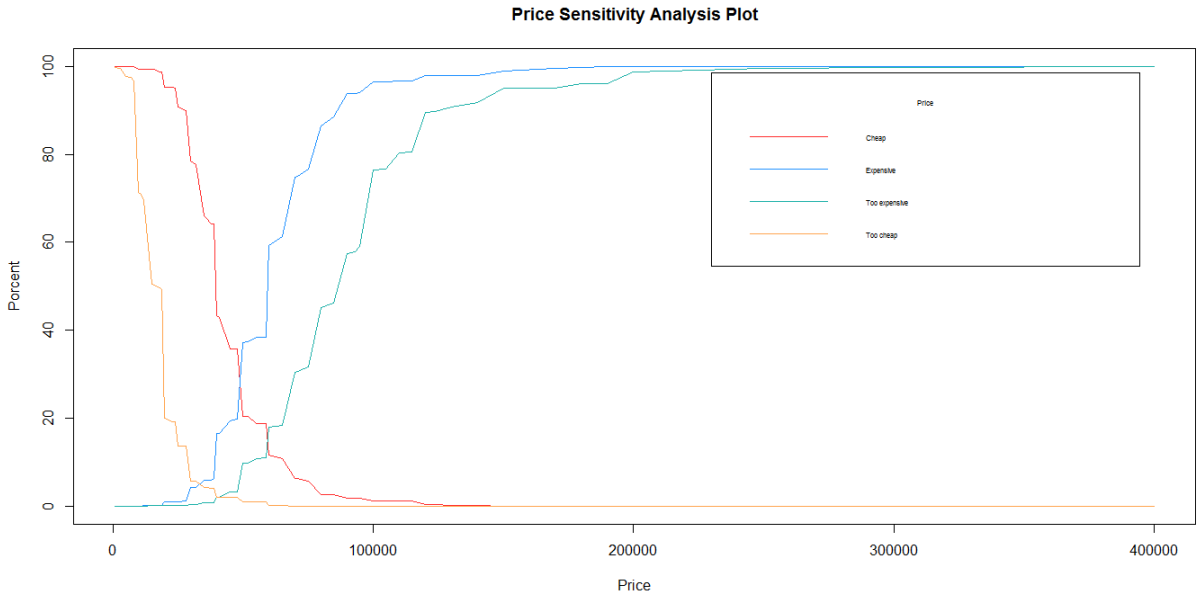


Gráfico 9: Gráfico Uso Modelo de Sensibilidad de Precios PSM (PSA)

Ahora bien, para ubicar los precios se usa la función locator() que se utiliza con el puntero del mouse sobre la intercepción OPP, IPP, LMB y LMC:

```
R> OPP <- locator()
R> OPP$x
```

Output: $x
[1] 41810.58

Donde muestra el valor para x que indica el precio donde está ubicado el punto para OPP.

R> IPP <- locator()

R> LMB <- locator()

R> LMC <- locator()

R> IPP$x

R> LMB$x

R> LMC$x

Output: IPP$x
[1] 48795.15
LMB$x
[1] 42509.04
LMC$x
[1] 49493.61

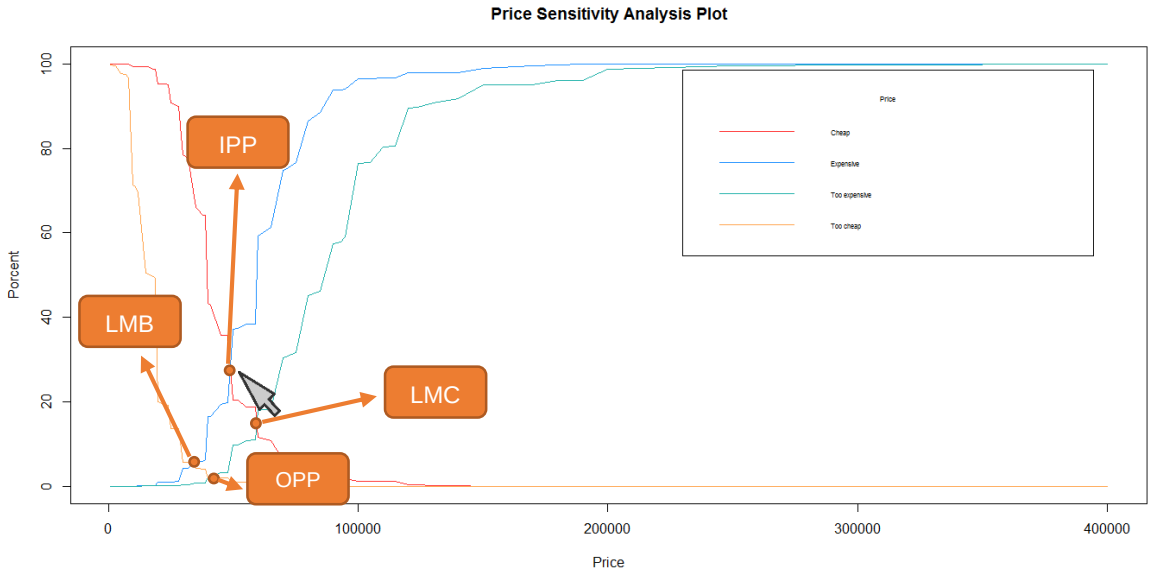


Gráfico 10: Identificación precios Uso Modelo de Sensibilidad de Precios PSM (PSA)

2.3. Uso Índice de Jaccard (Jac)

El uso de la función Jaccard describe el grado de asociación o semejanza entre los elementos comparados, expresado en valor numérico, entre 0 y 1, contiene los siguientes argumentos:

R> Jac(QG_prefix, A_prefix, data)

Donde:

QG_prefix	columnas que indagan por las marcas que satisfacen al cliente de forma general (usa "")
A_prefix	columnas que indagan por las preferencia de las marcas en un atributo (usa "")
data	dataframe que contiene las variables de interés

Tabla 33: Uso Índice de Jaccard (Jac)

Un ejemplo con un estudio de investigación de mercados sobre cuatro marcas de productos de belleza de venta directa, donde se simularon 100 casos a personas que deberían contestar las siguientes preguntas:

Z= ¿Cuáles de las siguientes marcas, cree usted que tienen los mejores productos de Belleza?,

1. Avon
2. Belcorp
3. Yanbal
4. Natura

A_1= ¿Cuáles de las siguientes marcas asociaría usted estos atributos, el empaque de los productos es el más atractivo?

1. Avon
2. Belcorp
3. Yanbal
4. Natura

A_2= ¿Cuáles de las siguientes marcas asociaría usted estos atributos, tiene unos precios competitivos en el mercado?

1. Avon
2. Belcorp
3. Yanbal
4. Natura

La base contiene una pregunta general *Z* para cada una de las marcas y dos atributos A_1 y A_2, que también incluyen cada una de las marcas:

ID	Z1	Z2	Z3	Z4	A1_1	A1_2	A1_3	A1_4	A2_1	A2_2	A2_3	A2_4
1		2	3	4		2	3				3	
2	1		3	4	1	2	3		1			4
3			3		1			4	1	2		4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	1	2		4	1						3	
99				4	1		3		1			4
100			3			2		4	1	2	3	

Cuadro 25: Ejemplo Uso Índice de Jaccard (Jac)

Como se puede apreciar en la base de datos el número del índice pertenece a la marca, utilizando la función Jac se realizara la asociación entre las respuestas de las marcas a la pregunta *Z (Pregunta General)* y un solo atributo A_1 *(Empaque Atractivo);* primero el dataframe utilizado “Base Jaccard.csv”, quedara almacenado en un objeto llamado Base:

```
R> Base <- read.table("Base Jaccard.csv", header=TRUE, sep=";", dec=",")
```

Luego se reemplaza los datos en la función con la pregunta *Z (Pregunta General)* y un solo atributo A_1 *(Empaque Atractivo):*

```
R> Jac(QG_prefix = "Z", A_prefix = "A1", data = Base)
```

Output:

```
[1] 0.395
```

La salida de la función es un dato que representa el valor de índice de Jaccard que es la asociación o semejanza entre las marcas comparadas, sin embargo un solo atributo no representa mucho, debe compararse con más de uno y de esta forma poder obtener un índice para cada uno de los atributos y dar respuesta a cual es más importante; se utilizara el mismo ejemplo de las marcas de productos de belleza, solo que se incluirá en simultanea los dos atributos al tiempo, A_1 *(Empaque Atractivo)* y A_2 *(Precios Competitivos),* para este caso se usa la función Jac_multiple:

R> Jac_multiple(QG_prefix, A_prefixes, data)

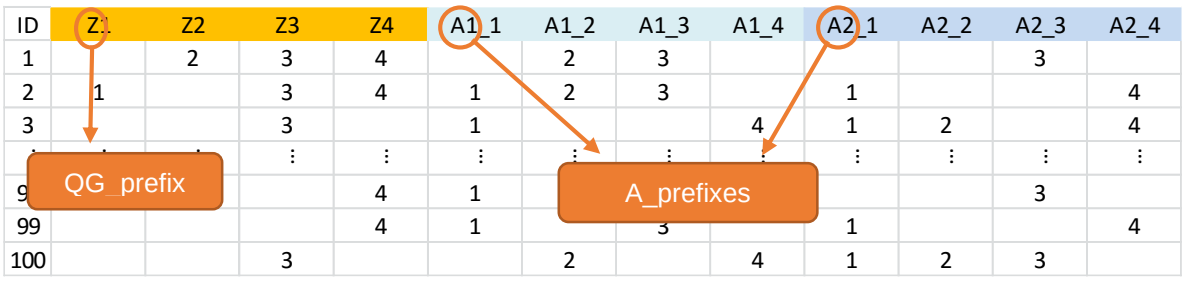
Donde:

QG_prefix	columnas que indagan por las marcas que satisfacen al cliente de forma general (usa "")
A_prefixes	lista de atributos que tengan las preferencias de las marcas (usa "")
data	dataframe que contiene las variables de interés

Tabla 34: Uso Índice de Jaccard (Jac_multiple)

El único cambio entre la función entre Jac y Jac_multiple es que se utiliza el argumento "A_prefixes" el cual pueden ingresarse los atributos por medio de una lista por ejemplo A_prefixes=c("A1","A2",...) .

Para identificar las variables que se deben usar en la función se observara la base:



ID	Z1	Z2	Z3	Z4	A1_1	A1_2	A1_3	A1_4	A2_1	A2_2	A2_3	A2_4
1		2	3	4		2	3				3	
2	1		3	4	1	2	3		1			4
3			3		1			4	1	2		4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
9				4	1						3	
99				4	1		3		1			4
100			3			2		4	1	2	3	

Cuadro 26: Identificación de Argumentos función (Jac_multiple)

De tal forma que se puede reemplazar los argumento de la función de la siguiente manera:

```
R> Jac_multiple(QG_prefix = "Z", A_prefixes = c("A1", "A2"), data = Base )
```

La salida muestra los dos resultados de los índices:

```
Output:
  Jac_A1    Jac_A2
0.3950000 0.4141667
```

El Jac_A1 (0.395) pertenece al primer atributo Empaque Atractivo y el valor del índice de Jac_A2 (0.4141), pertenece al atributo Precios Competitivos.

Como se indicaba anteriormente la principal razón de sacar los índices es obtener el más importante, de tal manera que con la siguiente instrucción puede organizarse jerárquicamente del mayor índice al menor:

```
R> cbind (Jaccard = salida2 [order(salida2, decreasing = T)])
```

De tal forma que la salida muestra cuál de los dos atributos tiene un índice mayor y puede decirse que tiene mayor importancia:

```
Output: Jaccard
Jac_A2 0.4141667
Jac_A1 0.3950000
```

Por lo que para los entrevistados que los productos de belleza de venta directa tengan Precios Competitivos es mayor a que los productos de belleza de venta directa tengan Empaques Atractivos.

2.4. Uso Penalty Analysis (PA)

El uso de la función de PA está enfocado en recomendar si se deben o no hacer ajustes a las variables organolépticas de un producto y además en qué sentido deben hacerse esos ajustes, los argumentos son los siguientes:

R> PA(GQ, JQ, data)

Donde:

GQ	pregunta general, puede contener una escala de agrado, satisfacción o intención de compra, debe tener cinco niveles de lo contrario debe recodificarse, el 1 con el menor nivel y el cinco con el mayor nivel (escala likert)
JQ	pregunta Just right o justo, debe contener una escala de tres niveles codificados de dos a cuatro donde tres será justo y dos y cuatro poco y mucho respectivamente
data	dataframe que contiene las variables de interés

Tabla 35: Uso Penalty Analysis (PA)

El siguiente ejemplo de un estudio de mercadeo que hace referencia a un helado de Crem Ice, marca que quiere descubrir si su producto tiene los aspectos adecuados para competir en el mercado, la base de datos contiene 100 casos simulados, donde las variables que lo componen son el ID (Identificador de la encuesta), una variable *G* (Intención de Compra) y un Atributo en este caso PJ (Nivel de dulce):

ID	G	PJ
1	2	2
2	5	3
3	1	2
4	3	3
⋮	⋮	⋮
99	4	4
100	1	4

Cuadro 27: Ejemplo Uso Penalty Analysis (PA)

$G =$ ¿Que tanto está interesado en comprar el producto que acaba de consumir, donde 1 es No me interesa compararlo y 5 me interesa mucho compararlo?

La escala de respuesta es de 5 puntos Likert es decir que satisface la condición del argumento, ahora bien la siguiente pregunta hace referencia a la escala JAR, que también tiene como requisito la función

$PJ =$ Para usted el nivel de dulce del helado que acabo de probar es…

1. Poco dulce para mi gusto

2. Justo como me gusta

3. Mucho dulce para mi gusto

La pregunta contiene 3 niveles, y el nivel de la mitad contiene la opción de respuesta Justo.

Identificados los argumentos que necesita la función se reemplaza según las variables de la base:

```
R> PA(GQ = G, JQ = PJ, data =Base)
```

La salida obtenida de la función es lista que contiene $penalty, $tukey_HSD y $summary, donde:

El valor $penalty contiene dos valores “PP” que indica el valor de Penalty Poco y “PM” valor de Penalty Mucho.

Output: $penalty

PP	PM
-0.5618182	-0.3563636

Según la siguiente tabla:

Menos a **(-0.25)** = bajo nivel: puede salir de este atributo solo
Entre **(-0,25)** y **(-0,50)** = nivel medio: considere hacer cambios a este atributo
Mayor a **(-0.50)** = alto nivel: definitivamente realizar cambios en este atributo

El valor de PP (-0.5618) tiene un nivel alto donde debe considerarse hacer cambios en el nivel de dulce Poco, es decir las personas compradoras del producto consideran que tiene poco dulce.

Ahora PM (-0.3563) tiene un nivel medio ya que se encuentran en el rango de entre (-0.25) y (-0.50), podría considerarse hacer cambios a Mucho dulce pero como el valor de PP es mayor se debe pensar primero hacer el ajuste a este nivel Poco.

La siguiente salida indica si la prueba tiene valides al considerase la prueba de HSD_tukey:

Output: $tukey_HSD

	diff	lwr	upr	p adj
PJ_3-PJ_2	1.7556818	1.1087472	2.4026164	1.29E-08
PJ_4-PJ_2	0.2708333	-0.4810809	1.0227476	6.68E-01
PJ_4-PJ_3	-1.4848485	-2.1914546	-0.7782424	7.52E-06

Los valores que se deben tener en cuenta son PJ_3-PJ_2 y PJ_4-PJ_3, donde se planteó la siguiente prueba de hipótesis:

- $H_0 = \overline{PJ}_3^1 = \overline{PJ}_2^1$ ó $2.6 = 4.3$

  $H_1 = \overline{PJ}_3^1 \neq \overline{PJ}_2^1$ ó $2.6 \neq 4.3$

Y

- $H_0 = \overline{PJ}_4^1 = \overline{PJ}_3^1$ ó $2.8 = 4.3$

  $H_1 = \overline{PJ}_4^1 \neq \overline{PJ}_3^1$ ó $2.8 \neq 4.3$

Para los dos casos existen evidencias estadísticas para rechazar H_0 es decir que existen diferencias significativas entre las medias.

Por último la salida:

Output: $summary

	JQ	n	f	avg_GQ
1	2	32	0.32	2.562500
2	3	44	0.44	4.318182
3	4	24	0.24	2.833333

Da una tabla de resumen con los datos involucrados en el análisis, JQ es el nivel de la variable *PJ,* n es las frecuencias absolutas para cada nivel y f la frecuencias relativas y por ultimo avg_GQ indica la media correspondiente a cada nivel de JQ.

De esta última tabla pude graficarse las frecuencias relativas de la tabla:

```
R> barplot(analisis1$summary$f,main="Penalty Analysis",
+ names.arg=c("Poco", "Justo", "Mucho"), col=c("darkblue","red","green"))
```

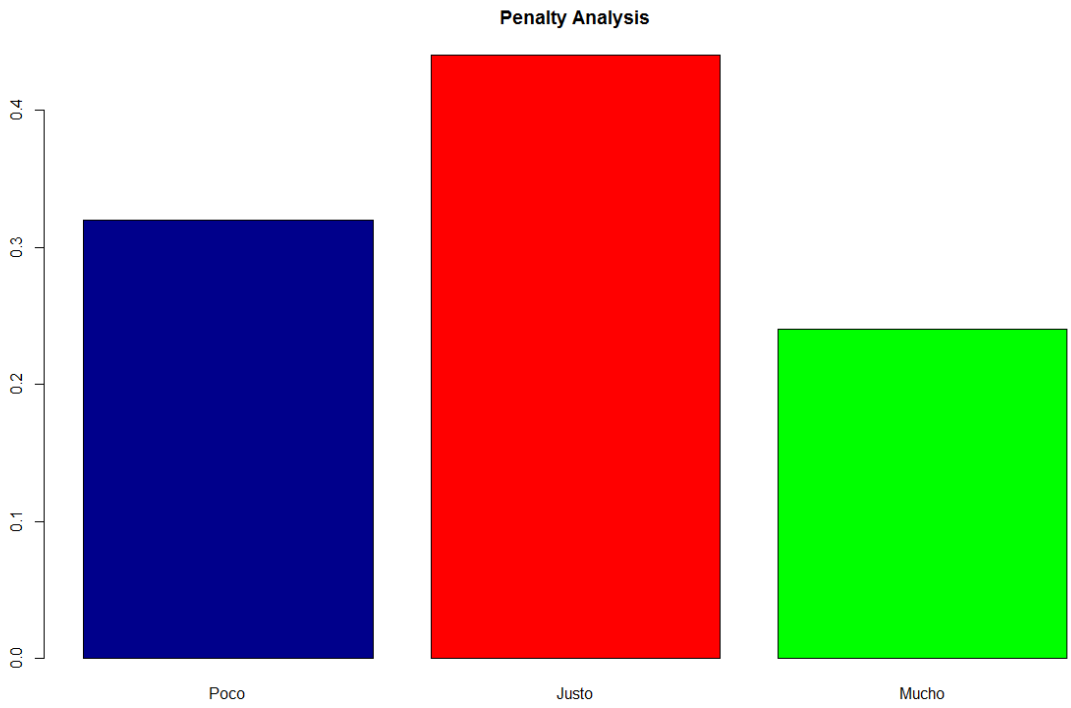


Gráfico 11: Gráfico Uso Penalty Analysis (PA)

En la gráfica se puede apreciar que la barra de Poco es mayor que la de Mucho por lo que a simple vista podría notarse que Poco debe tener un ajuste en el nivel de dulce.

CONCLUSIONES

Se solucionaron los problemas típicos mostrados a los largo del documento y que presentan información estadística que es utilizada en estudios de investigación de mercados.

Se realizaron herramientas computacionales que sirven para solucionar los problemas más frecuentes en la depuración de bases de datos, estos problemas abordados fueron: inconsistencia de manejo de respuestas múltiples, problemas de coherencia entre conjuntos de variables múltiples, duplicidad de información, y problemas en las validaciones de datos en los análisis de sensibilidad de precios, en los cuales se identificó cada una de las fragilidades que presentaba cada problema y se ilustro de forma que fuera fácil encontrar las encuestas que presentaba contrariedades y poder dar la solución más apropiada a cada uno de los ejemplos mostrados.

También se realizaron rutinas en software estadístico que permitieron implementar análisis que usualmente se usan en estudios de investigación de mercados, como el Modelo Kano, el cálculo del Índice de Jaccard, Análisis de Penalidades (conocido en la literatura anglosajona como penalty analysis), y el análisis de sensibilidad de Precios de Van Westendorp (en inglés Van Westendorp's Price Sensitivity Meter - PSM), para cada uno de estos se generó una salida (output) correspondiente a cada análisis y se realizó ejemplos prácticos que permiten por medio de gráficos cuando lo ameritan ilustrar mejor cada uno de los análisis.

REFERENCIAS

Fuentes de datos: Datos simulados

BIBLIOGRAFÍA

American Marketing Association. (2004, Octubre). *https://www.ama.org*. Retrieved from https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx

BS, L. (24 de Abril de 2009). *https://es.scribd.com/*. Obtenido de https://es.scribd.com/doc/14590547/11/Prueba-de-Tukey

Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T.-J. (s.f.). *http://viceroy.eeb.uconn.edu/*. Obtenido de http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstSUsersGuide/References/ChaoEtAl2005Sp.pdf

Chatterjee, S., Singh, N., Goyal, D., & Gupta, N. (2014). *Managing in Recovering Markets.* India.

Goodpasture, J. C. (2004). *Quantitative Methods in Project Management.* Estados Unidos.

Guzmán Salas, Z. (21 de Noviembre de 2008). *http://catarina.udlap.mx*. Obtenido de http://catarina.udlap.mx/u_dl_a/tales/documentos/lhr/guzman_s_z/capitulo4.pdf

ICC / ESOMAR International Code on Market and Social. (2007). *https://www.esomar.org*. Retrieved from https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR_Code_English_.pdf

Justel, A. (s.f.). *https://www.uam.es*. Obtenido de https://www.uam.es/personal_pdi/ciencias/ajustel/docencia/ad/AD10_11_Cluster.pdf

León Duarte, J. A. (2005). *Tesis Doctotales en Red TDR.* Recuperado el 01 de Mayo de 2015, de http://www.tdx.cat/bitstream/handle/10803/6840/03Jld03de08.pdf?sequence=3

Ludwig, J. A., & Reynolds, J. F. (1988). *Statistical Ecology: A Primer in Methods and Computing.* Estados Unidos.

Pati, D. (2002). *Marketing Research.* India.

Pereira, J. E. (28 de Enero de 2010). *http://www.mercadeo.com*. Obtenido de http://www.mercadeo.com/blog/2010/01/investigacion-de-mercados/

Sáez Castillo, A. J. (8 de Julio de 2008). *https://cran.r-project.org*. Obtenido de https://cran.r-project.org/doc/contrib/Saez-Castillo-RRCmdrv21.pdf

Sarstedt, M., & Mooi, E. (2014). *A Concise Guide to Market Research.* Springer-Verlag Berlin Heidelberg.

The Center for Quality of Management Journal. (Junio de 1993). *http://www.clientservice.ru*. Obtenido de http://www.clientservice.ru/wp-content/uploads/2007/06/kano-model-in-use.pdf

Trespalacios Gutiérrez, J. A., Vázquez Casielles, R., & Bello Acebrón, L. (2005). *Investigación de mercados: métodos de recogida y análisis de la información para la toma de decisiones en marketing.* España: Ediciones Paraninfo.

unac.edu.pe. (Abril de 2011). *http://www.unac.edu.pe/*. Obtenido de http://www.unac.edu.pe/documentos/organizacion/vri/cdcitra/Informes_Finales_In

vestigacion/Abril_2011/IF_VIVANCO_FIPA/l_Capitulo%208_Pruebas%20estad%EDsticas.pdf

Varela, P., & Ares, G. (2014). *Novel Techniques in Sensory Characterization and Consumer Profiling.*

Walker, L. (s.f.). *http://www.fona.com*. Obtenido de http://www.fona.com/sites/default/files/WhitePaper_Penalty%20Analysis_sensory.pdf

Witell, L., Lofgren, M., & Dahlgaard, J. J. (2013). *http://liu.diva-portal.org*. Obtenido de http://liu.diva-portal.org/smash/get/diva2:666390/FULLTEXT02.pdf