

APLICACIÓN DE UN DISEÑO DE MEDIDAS REPETIDAS PARA EXPLICAR SI
EL TIEMPO ES UN FACTOR DETERMINANTE EN LAS VENTAS DE
PRODUCTOS DE CONSUMO MASIVO, DURANTE EL PERIODO
COMPRENDIDO ENTRE LOS AÑOS 2013 Y 2014

LUZ AMANDA PACHECO SOLANO

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BASICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C., 2017

APLICACIÓN DE UN DISEÑO DE MEDIDAS REPETIDAS PARA EXPLICAR SI
EL TIEMPO ES UN FACTOR DETERMINANTE EN LAS VENTAS DE
PRODUCTOS DE CONSUMO MASIVO, DURANTE EL PERIODO
COMPRENDIDO ENTRE LOS AÑOS 2013 Y 2014

LUZ AMANDA PACHECO SOLANO

Trabajo de grado realizado para obtener el título de especialistas en Estadística
Aplicada

Wilmer Pineda Ríos
Magister en Matemáticas
Director de trabajo de grado

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BASICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C., 2017

Nota de aceptación

Presidente del Jurado

Jurado:

Jurado:

Bogotá, D. C Febrero de 2017

Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documentos. Estos corresponden únicamente a los autores.

AGRADECIMIENTOS

Doy gracias a mi familia por ayudarme a darle continuidad a este proceso, a mi esposo que me impulso a continuar y a no desistir, a mi hijo que es el motor que me estimula a salir adelante y a quien quiero darle el mejor ejemplo de persistir y no desistir.

CONTENIDO

1	INTRODUCCIÓN.....	12
2	LA EMPRESA.....	13
3	MARCO TEÓRICO.....	17
3.1	MERCADO DE CONSUMO MASIVO.....	17
3.2	OFERTA Y DEMANDA.....	18
3.2.1	Demanda.....	18
3.2.2	Oferta.....	18
3.3	VARIABLES ECONÓMICAS.....	19
3.3.1	Descripción de variables.....	20
4	MEDIDAS REPETIDAS.....	21
4.1	CARACTERÍSTICAS GENERALES DEL DISEÑO DE MEDIDAS REPETIDAS.....	21
4.2	EL ANÁLISIS DE DATOS EN LOS DISEÑOS DE MEDIDAS TOTALMENTE REPETIDAS.....	22
4.2.1	Supuestos básicos para el análisis y alternativas ante su incumplimiento.....	22
4.3	ASPECTOS DE DISEÑO.....	24
4.4	EL ENFOQUE UNIVARIADO.....	24
4.4.1	Supuestos.....	24
4.5	EL ENFOQUE MULTIVARIADO.....	25
4.5.1	Supuestos.....	25
5	ANÁLISIS DE DATOS.....	27
5.1	ANÁLISIS DESCRIPTIVO DE LA MUESTRA.....	28
5.2	MUESTREO.....	29
5.2.1	Requisitos de una buena muestra.....	29
5.2.2	Muestreo Aleatorio Simple.....	30
5.2.3	Muestreo Estratificado.....	31
5.3	ANÁLISIS POR CANAL DE DISTRIBUCIÓN.....	35
5.4	MODELO.....	37
6	CONCLUSIONES.....	42
	BIBLIOGRAFÍA.....	44

LISTA DE TABLAS

Tabla 1. Análisis Descriptivo General de Empresa (ventas en MM)	27
Tabla 2. Muestreo Aleatorio Simple Estratificado (EST-MAS)	34
Tabla 3. Promedio por Canal de Distribución	35
Tabla 4. Estadísticos por Canal de Distribución	35
Tabla 5. Estadísticos por Grupo de Clientes.....	36

LISTA DE GRÁFICAS

Gráfica 1. Ventas anuales de la Empresa.....	13
Gráfica 2. Ventas Acumuladas año 2013 y 2014	14
Gráfica 3. Ventas Mensuales por Canales.....	15
Gráfica 4. Participación Porcentual por Canales.....	15
Gráfica 5. Análisis Descriptivo de la Muestra (Participación de Clientes por Canales)	28
Gráfica 6. Comportamiento de los canales a través del tiempo	38
Gráfica 7. Gráfico de cajas y bigotes de las Ventas Vs Canal de Distribución	39
Gráfica 8. Gráfico de cajas y bigotes de las Ventas Vs Periodos de Venta	40
Gráfica 9. Gráfico de cajas y bigotes de las Ventas por Canal Vs Periodos de Venta año 2013.....	40
Gráfica 10. Gráfico de cajas y bigotes de las Ventas por Canal Vs Periodos de Venta año 2014.....	41

GLOSARIO

CANAL DE DISTRIBUCIÓN: Es el circuito mediante el cual la Empresa pone a disposición los productos para el consumidor final.

CONSUMO: Se conoce al resultado de consumir (verbo que se utiliza cuando se hace uso de un bien o servicio, o cuando se busca un sinónimo de gastar). Utilizar comestibles u otros bienes para satisfacer necesidades o deseos.

CONSUMIDOR: Individuo que consume bienes y servicios en el mercado.

CRECIMIENTO SOSTENIDO: Es aquel crecimiento que se manifiesta durante un largo periodo de tiempo.

ESTACIONALIDAD: Es aquel comportamiento cíclico de los datos que se mantiene en diferentes periodos de tiempo.

ETARIA: Pertenciente o relativo a la edad de una persona. Dicho de varias personas: Que tienen la misma edad.

MARCO DE MUESTREO: Es la lista de unidades de muestreo. Para encuestas telefónicas, el marco de muestreo podría ser una lista de todos los números de teléfono residencial de la ciudad; para entrevistas personales, una lista de todas las direcciones de calle; para una encuesta agrícola, una lista de todas las granjas o un mapa de las áreas que contienen granjas.

MUESTRA: Un subconjunto de una población.

MUESTRA POBLACIONAL: La colección de todas las posibles unidades de observación que podría haber sido elegidos en una muestra; la población de la que se tomó la muestra.

MULTINACIONAL: Empresa que tiene negocios y actividades establecidos en varios países.

PIRÁMIDE DE MASLOW: Es una teoría psicológica que inquiriere acerca de la motivación humana. Según Abraham Maslow, un psicólogo humanista, nuestras acciones están motivadas para cubrir ciertas necesidades.

POBLACIÓN OBJETIVO: Es población completa de observaciones que deseamos estudiar. Una parte importante y a menudo difícil del estudio es definir la población objetivo. La elección de la población objetivo afectará profundamente las estadísticas que resultan.

UNIDADES DE NEGOCIO: Diferentes negocios dentro de una misma empresa.

UNIDAD DE OBSERVACIÓN: Un objeto en el que se realiza una medición. Esta es la unidad básica de observación. A veces llamado elemento. En el estudio de las poblaciones humanas, unidades de observación son a menudo los individuos.

UNIDAD DE MUESTREO: Se refiere a la unidad que probamos en realidad. Se puede querer estudiar los individuos pero no se tiene una lista de todos los individuos de la población objetivo. En cambio, los hogares sirven como las unidades de muestreo, y las unidades de observación son las personas que viven en los hogares.

VARIABLE ECONÓMICA: Nombre que puede adoptar cualquiera de los agregados macroeconómicos tales como inversión, consumo, ahorro, gasto, ventas, etc., para su análisis en el tiempo y en el espacio.

VARIABLE CONTINUA: Una variable continua es aquella que puede adoptar cualquier valor en el marco de un intervalo que ya está predeterminado.

VARIABLE DISCRETA: Es aquella que solo puede adquirir un valor de un conjunto de números.

VARIABLE CUALITATIVA: Es aquella que presenta modalidades no numéricas que no admiten un criterio de orden.

RESUMEN

Este trabajo se enfoca en demostrar la influencia del tiempo en las ventas durante los años 2013 y 2014 y, para ello se utilizó el diseño de medidas repetidas como herramienta de análisis que permitiera desarrollar un modelo donde las ventas, el canal de distribución y el tiempo fueran las variables de estudio. De igual manera, se utilizaron otras herramientas estadísticas como los métodos de muestreo, que permitieran tomar una muestra de la población con la cual hacer el análisis correspondiente. Para ello, se ahondó en la teoría del Muestreo Aleatorio Simple y Muestreo Estratificado. Se hizo necesario hacer un muestreo estratificado debido a los diferentes canales de distribución por los cuales se entrega el producto al consumidor final. Las herramientas estadísticas aplicadas y las estimaciones permitieron evidenciar que el tiempo es un factor importante para pronosticar.

Palabras clave: *medidas repetidas, muestreo aleatorio simple, muestreo estratificado, modelo.*

ABSTRACT

This work focuses on demonstrating the influence of time on sales for the years 2013 and 2014 and to this end the repeated measures design analysis tool that would develop a model was used where sales, distribution channel and time were the study variables. Similarly, other statistical tools such as sampling methods, which allow a sample of the population with which to make the corresponding analysis were used. To do this, he delved into the theory of Simple Random Sampling and Stratified Sampling. It became necessary to make a stratified sampling due to the different distribution channels through which the product is delivered to the final consumer. The statistics tools applied and allowed estimates show that time is an important factor to predict.

Keywords: *repeated measures , simple random sampling , stratified sampling, model.*

1 INTRODUCCIÓN

Siendo el mercado el lugar en donde se encuentran demandantes y oferentes para realizar transacciones comerciales, es obvio encontrar empresas que se dedican a comercializar diferentes tipos de productos, entre ellos productos de consumo masivo. La incertidumbre de la economía dado el libre mercado y la libre competencia, empujan a las empresas a crear estrategias de mercadeo y ventas, diferentes tipos de negociación y alianzas que permitan llegar a los consumidores a través de sus clientes directos.

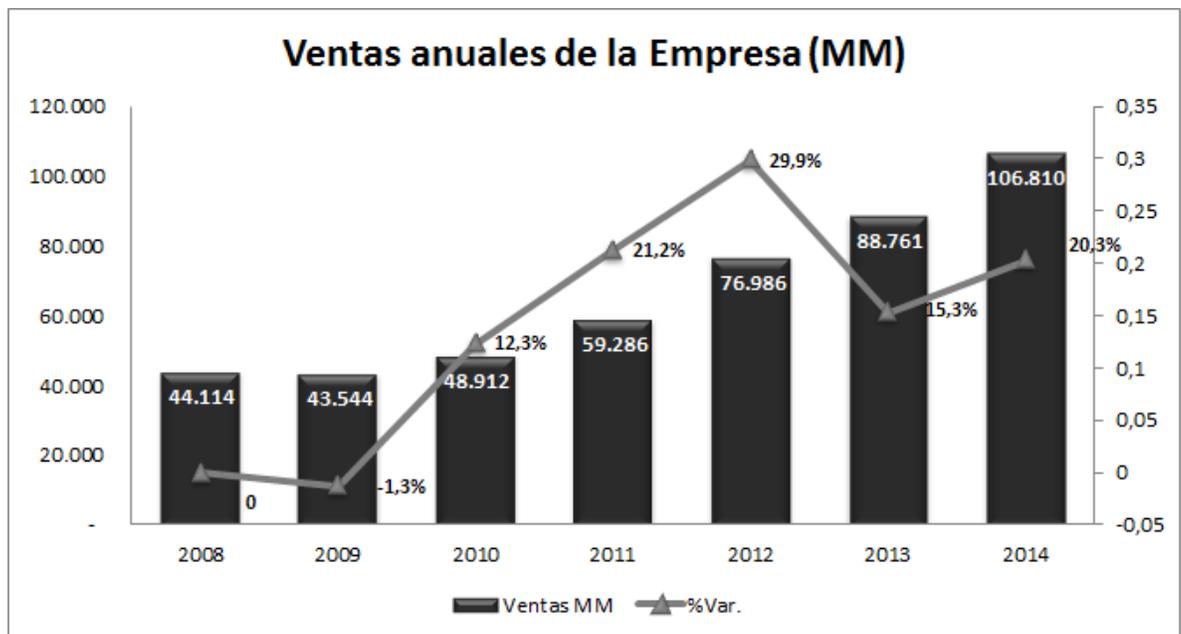
Todos estos comportamientos hacen que las empresas diariamente se enfrenten a la incertidumbre del mercado y muchas veces deben enfrentarse a la toma de decisiones inmediatas que las lleven a cumplir con los objetivos propuestos en los diferentes ámbitos financieros, productivos, logísticos, etc., siendo su principal objetivo ganar más espacio en el mercado, alcanzar los mejores indicadores de desempeño, con los cuales tener mayor utilidad, entre otros, y para ello es importante poder contar con herramientas de análisis que les muestren a la mayor brevedad el comportamiento de sus datos y que estos indiquen en dónde deben enfocar sus esfuerzos para alcanzar la meta propuesta.

Dado lo anterior, las empresas están constantemente generando estadísticas basadas en los datos recolectados de las actividades diarias. Sin embargo, tener datos bien analizados bajo modelos estadísticos que hayan sido puestos a prueba con parámetros de medición es lo que toda empresa espera y por ello en este trabajo se diseñará un modelo de medidas repetidas, que permita identificar si efectivamente el tiempo es factor determinante para las ventas.

En el presente documento el lector encontrará, en primer lugar, un análisis de los datos históricos con lo cual se hizo un primer acercamiento al comportamiento de la Empresa. En segundo lugar, encontrará el procedimiento de extracción de la muestra de acuerdo a la estructura comercial de la Empresa, previa determinación de las variables económicas que harán parte del modelo. Por último, se aplicó la teoría de Medidas Repetidas y con la ayuda de un software estadístico se obtuvo el modelo correspondiente.

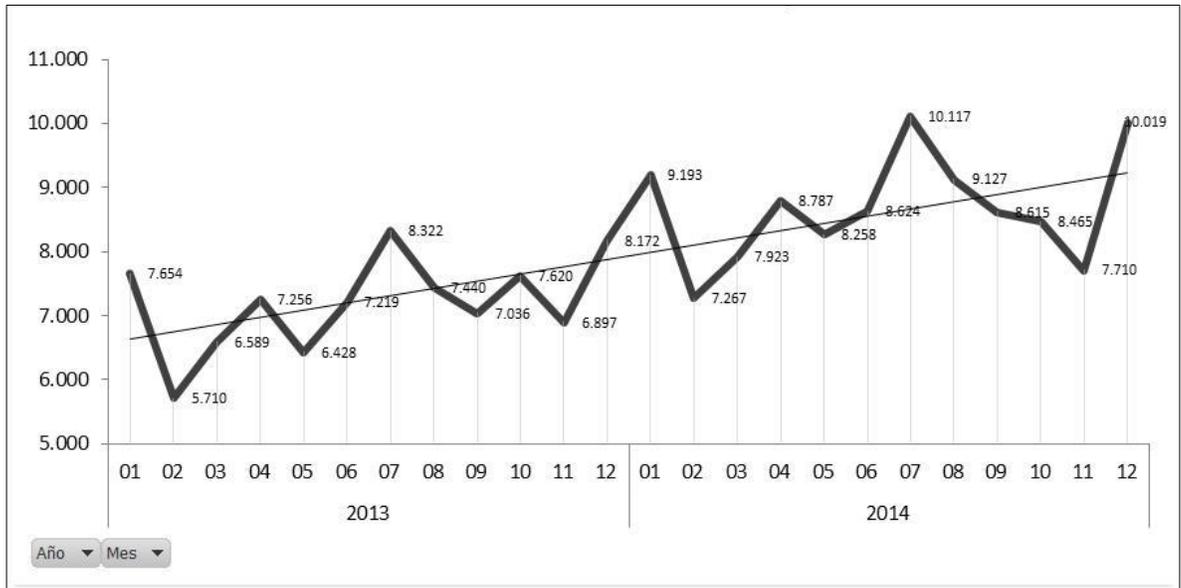
2 LA EMPRESA

La Empresa estudio de este trabajo pertenece a la industria de comestibles (consumo masivo) con más de 50 años de experiencia en el mercado. Desde su creación, la Empresa ha tenido una constante evolución, que parte de los productos que produce y comercializa, pasando por la apertura de diferentes canales de distribución y llegando al gusto de diversos consumidores que finalmente redonda en los ingresos percibidos. La Empresa ha venido presentando un crecimiento sostenido durante los últimos cinco años como se puede apreciar en la figura 1, donde se muestra su evolución anual.



Gráfica 1. Ventas anuales de la Empresa

En la figura 2 se observa una tendencia creciente de las ventas, manteniéndose durante los mismos periodos de ambos años un comportamiento similar, es decir, mostrando picos en los mismos meses, lo que indica que existe estacionalidad en las ventas.

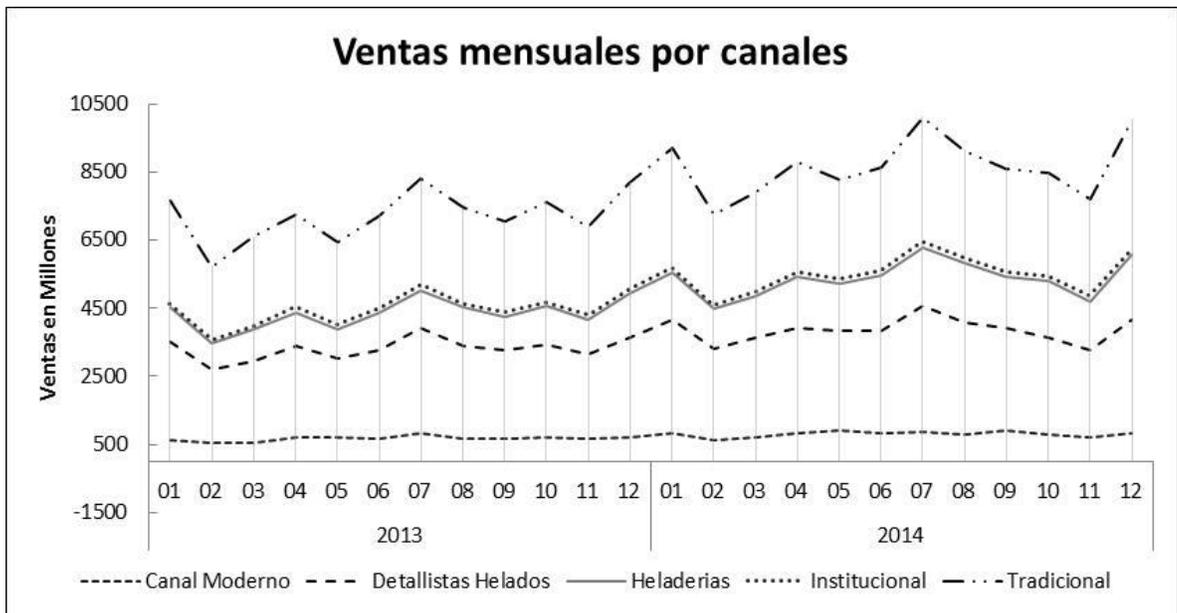


Gráfica 2. Ventas Acumuladas año 2013 y 2014

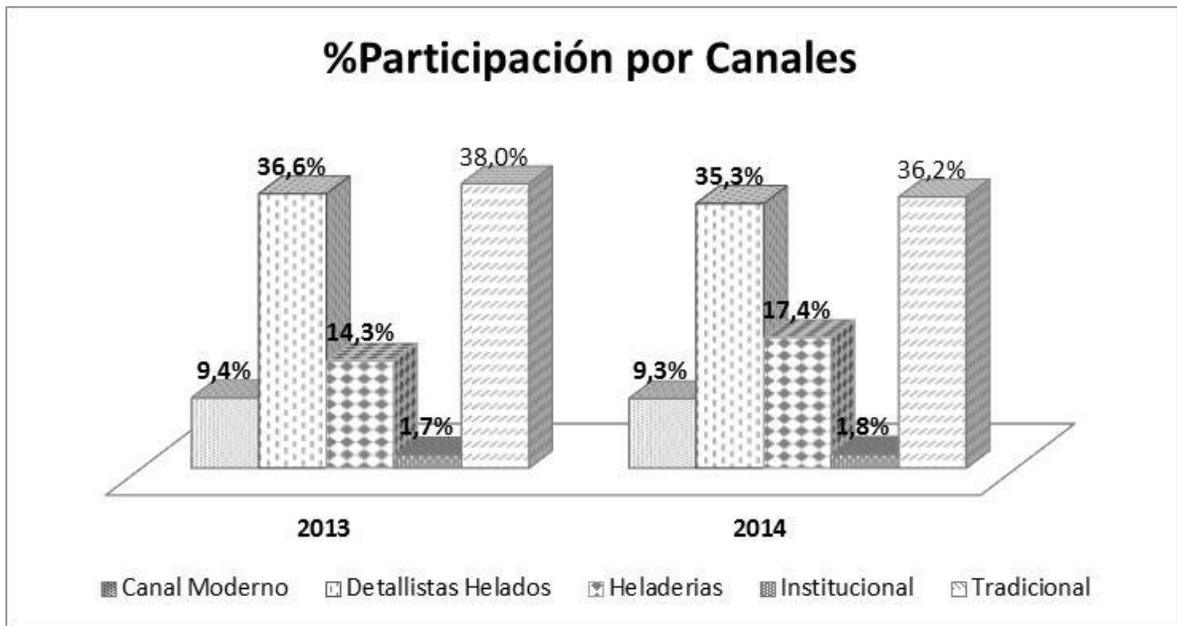
Para comercializar los productos de la Empresa, es necesario contar con una infraestructura apta para el almacenamiento y conservación de los mismos. Por ello la Empresa cuenta con una flota de camiones propios con los cuales distribuye sus productos diariamente en diferentes ciudades del territorio colombiano, incluyendo ciudades como Bogotá, Bucaramanga, Cali, Medellín, Barranquilla, Pereira, Montería, entre otras. Igualmente, existen poblaciones a las cuales la Empresa no llega directamente por temas de costos y de acceso, sino que lo hace a través de empresas distribuidoras de sus productos.

La Empresa ha logrado incursionar en diferentes canales de distribución como tienda a tienda (TAT), almacenes de cadena, heladerías fruterías, food service, instituciones educativas, entre otros tipos de negocio.

En la figura 3 se puede observar que todos los canales tienen un comportamiento similar en los diferentes meses, mostrando una tendencia uniforme. Sin embargo, en la figura 4 se puede determinar que el canal que más participa en las ventas es el canal tradicional, correspondiente a las ventas de mayor volumen (distribuidores), seguido de detallistas helados que corresponde a la venta tienda a tienda. Igualmente se observa que el comportamiento es similar en ambos años.



Gráfica 3. Ventas Mensuales por Canales



Gráfica 4. Participación Porcentual por Canales

Siendo la Empresa una multinacional que se expande rápidamente, y la cual está dividida en varias unidades de negocios, requiere identificar de forma rápida aquellas variables críticas que son fundamentales para su buen desempeño, y que trazan el horizonte hacia el cual la Empresa se debe encaminar, y que a la menor variación se puedan tomar acciones rápidas para afrontar los cambios que se puedan presentar en el mercado.

En la actualidad la Empresa cuenta con gran cantidad de información comercial que de manera individual es manejada por las unidades de negocio y que finalmente se suman en un gran total. Sin embargo, para que ese gran total sea el esperado por los accionistas se debe analizar puntualmente las cifras tanto en pesos como en unidades de venta, para identificar aquellas oportunidades con las que se puede llegar al nivel esperado.

Muchas veces el tener gran cantidad de información no garantiza que la Empresa tenga las herramientas para aprovechar las oportunidades del mercado y por el contrario puede llegar a caer en la generación y almacenamiento innecesario de datos que al final no aportan nada para alcanzar los indicadores esperados.

En este caso particular, no se cuenta con un modelo en el cual apoyarse para tomar decisiones de comercialización, teniendo que acudir a ejercicios básicos de participación de las variables, que muchas veces esconden el verdadero comportamiento de los datos. Con la aplicación de este modelo se espera mejorar en cuanto a procedimiento de análisis de datos basándose en herramientas estadísticas sólidas que permitan mostrar una mejor consolidación de la información.

3 MARCO TEÓRICO

3.1 MERCADO DE CONSUMO MASIVO

Los mercados de consumo masivo se identifican por tener algunas características específicas. Bajo los criterios descritos a continuación se encuentran clasificados los productos de la Empresa en estudio. Como su nombre lo dice, el acceso a los productos comercializados en este segmento debe ser masivo, es decir, que una gran masa de personas puede y debe tener acceso a la compra de los mismos. Al ser masivo, el producto no distingue en ser demandado por franja etaria, geográfica, social, (un niño o un adulto demandan lo mismo, como así también una mujer o un hombre), salvo raras y escasas excepciones que tienen que ver con necesidades muy específicas como las toallas femeninas o preservativos para los hombres.

Cuando un producto, está en esta clasificación por lo general se considera la “P” de Plaza como la más importante, debido a que si el consumidor no encuentra el producto en determinado local este no tendrá inconveniente de buscar en otro punto de venta.

Por lo citado anteriormente, es que las empresas que comercializan sus productos en este tipo de mercado, apuestan la mayoría de sus esfuerzos en la “P” de Plaza, esta decisión también puede ser determinante para descifrar en qué mercado se encuentra una empresa.

Finalmente el criterio más importante para determinar si un mercado es de consumo masivo, es analizar si en este mercado se cumplen alguna de las necesidades descritas en la Pirámide Maslow (alimentación/bebida, descanso, higiene personal, comunicación, transporte). En esta clasificación los consumidores no están dispuestos en hacer grandes esfuerzos en sus compras, debido a que no pasan por muchos procesos de análisis de los productos, no identifican con facilidad las diferencias que tienen los productos ofertados. El proceso de compra es mayormente simple, debido a que el monto a desembolsar no es mucho y por ende el proceso carece de complejidad.

El mercado de consumo masivo se divide en varios conceptos:

- Segmento de Consumo Masivo Básico.
- Segmento de Consumo Masivo de Urgencia.
- Segmento de Consumo Masivo por Impulso.
- Segmento de Consumo Masivo de Preferencia.

De acuerdo a la clasificación anterior los productos vendidos en la Empresa se agrupan en el segmento de consumo masivo por impulso, el cual es un mercado que está prácticamente guiado por los estímulos visuales donde las compras suelen carecer de racionalidad, apoyado también por la no planificación ni búsqueda de la compra. Básicamente se relacionan con la autoestima, el antojo y el deleite.¹

3.2 OFERTA Y DEMANDA

3.2.1 Demanda

La cantidad de un artículo que los consumidores desean adquirir varía en función de factores como el precio del artículo, el ingreso económico del consumidor, los precios de los demás artículos y la preferencia, que es de carácter subjetivo.

La curva de demanda indica cuánto están dispuestos a comprar los consumidores de un bien cuando varía el precio unitario.

Igualmente se encuentra la demanda temporal la cual se refiere a los niveles de demanda comunes en los pronósticos. La variación de la demanda en el tiempo es resultado del crecimiento o declinación de los índices de ventas, variación estacional del patrón de demanda, así como de las fluctuaciones generales ocasionadas por múltiples factores.

3.2.2 Oferta

La cantidad de un artículo dado que un productor individual está dispuesto a vender en un cierto período de tiempo, depende del precio del artículo y de los costos de producción.

La oferta de determinado artículo se define como “las diferentes cantidades que los productores estarán dispuestos y en condiciones de ofrecer en el mercado en función de los distintos niveles de precios posibles, en determinado tiempo.”

La curva de oferta muestra la cantidad que están dispuestos los productores a vender a de un bien a un precio dado, manteniendo constantes los demás factores que pueden afectar la cantidad ofrecida.

¹ VÁZQUEZ, Diego, Mercados de Consumo, 2014, <http://www.mktdiegovazquez.com/2011/02/mercados-de-consumo.html> [Consulta: Jueves, 13 de Agosto de 2015].

3.3 VARIABLES ECONÓMICAS

La economía es la ciencia que estudia los recursos, la creación de riqueza y la producción, distribución y consumo de bienes y servicios, para satisfacer las necesidades humanas.

Igualmente, se encarga de administrar los bienes escasos, por lo cual se enfoca en estudiar el comportamiento de variables y la relación existente entre ellas.

Una variable económica es la representación de un concepto económico que puede medirse o tomar diversos valores numéricos. Se pueden encontrar variables cualitativas y cuantitativas.

El dinamismo del mercado está enmarcado entre la oferta y la demanda, por lo cual se debe tener claridad de estos dos conceptos:

Se llama cantidad demandada de un bien a la cantidad de ese bien que las unidades de consumo (familias) desean comprar. La demanda es igual a comprar y no es otra cosa que la cantidad de bienes que pueden ser consumidos en el mercado por la economía familiar de los compradores y está en función directa a su capacidad de compra y al nivel de ingresos que dispone la economía familiar si se incrementa el precio de un producto automáticamente disminuye su demanda y el consumidor procura sustituir su consumo con otro bien cuyo precio sea más conveniente y tenga condiciones similares en la satisfacción de sus necesidades.

La función oferta entendiéndose como la venta en el mercado no es otra cosa sino la capacidad que tiene los productores y vendedores de entregar al mercado bienes y servicios los cuales tienen una relación directa entre el precio de un bien y su cantidad de oferta por lo tanto si se incrementa el precio se incrementa la cantidad ofertada y al contrario si se produce una disminución o caída de precios se tiene un decrecimiento de la cantidad ofertada. Entre las variables que determinan la oferta están el precio del bien, precios de otros bienes, costos de los factores de producción, tecnología, invención, innovación, preferencias o gustos del consumidor.

Entre otras variables que inciden tanto en la demanda como en la oferta, se pueden observar el tiempo, el clima, estructuras empresariales, canales de distribución, la innovación, entre otros.

En el presente trabajo se dispondrá de cuatro variables que se analizarán con base en el tiempo, para determinar si éste repercute en el resultado de las ventas.

3.3.1 Descripción de variables.

Para el análisis de los datos disponibles se consideraron cuatro variables:

VARIABLE	DESCRIPCIÓN	TIPO
Ventas (Ingreso)	Corresponde al ingreso mensual por efecto de vender un producto a un precio determinado.	Continua
Canal de distribución	Hace referencia a la clasificación dentro de la empresa que se le da a los diferentes tipos de clientes. Este grupo de clientes define el canal de distribución a través del cual será atendido.	Cualitativa (Nominal)
Tiempo (meses)	Lapso en el cual se han realizado transacciones comerciales con los clientes.	Discreta

4 MEDIDAS REPETIDAS

4.1 CARACTERÍSTICAS GENERALES DEL DISEÑO DE MEDIDAS REPETIDAS

Los diseños de medidas repetidas, conocidos también como diseños intrasujeto y diseños de medida múltiple, se caracterizan por el registro de diversas medidas de la variable dependiente en un mismo grupo de sujetos. De este modo, las comparaciones entre las respuestas de los sujetos, ante los distintos tratamientos, se llevan a cabo dentro de un único grupo de sujetos (comparaciones intrasujeto), no estableciéndose comparaciones entre diferentes grupos de sujetos (comparaciones intersujetos o intergrupos).²

El interés por el diseño intrasujeto radica en la posibilidad que éste brinda de registrar un conjunto de puntuaciones o medidas de una variable, en dos o más puntos en el tiempo. Así, dado que desde una perspectiva longitudinal las respuestas de cada uno de los sujetos son función del tiempo, el diseño de medidas repetidas constituye un instrumento muy útil para la modelización de las curvas de crecimiento y la evaluación de los procesos de cambio en contextos evolutivos, sociales y educativos.

Los diseños de medidas repetidas presentan varias ventajas con respecto a los diseños de medida única, a saber:

- En primer lugar, constituyen instrumentos excelentes para reducir la varianza de error y minimizar la varianza sistemática secundaria. Dado que, en la estrategia intrasujeto el propio sujeto se convierte en criterio de bloqueo o de control, se extrae de la variabilidad del error una de sus principales fuentes, a saber, la varianza procedente de las diferencias individuales. De esta manera, el diseño intrasujeto consigue mayor precisión que cualquier otro tipo de diseño en la estimación de los efectos experimentales. De acuerdo con esta cualidad, diversos autores afirman que los diseños de medidas repetidas, poseen mayor potencia estadística que los diseños completamente aleatorios.
- Una segunda ventaja que tiene gran importancia, radica en la menor cantidad de sujetos que se requieren para llevar a cabo el estudio con

² BALLUERKA LASA, Nekane y VERGARA IRAETA, Ana Isabel. Diseños de Investigación Experimental en Psicología. Pearson Educación S.A. p. 241.

respecto a las investigaciones en las que se utilizan diseños de grupos totalmente al azar.

4.2 EL ANÁLISIS DE DATOS EN LOS DISEÑOS DE MEDIDAS TOTALMENTE REPETIDAS

4.2.1 Supuestos básicos para el análisis y alternativas ante su incumplimiento.

El modelo analítico utilizado habitualmente para llevar a cabo la prueba de hipótesis con este tipo de diseños se conoce como análisis de la varianza mixto. Si consideramos el formato más simple del diseño de medidas repetidas, el modelo matemático no aditivo que subyace al análisis de la varianza, bajo el supuesto de la hipótesis alternativa responde a la siguiente expresión:

$$y_{ij} = \mu + \eta_i + \alpha_j + (\eta\alpha)_{ij} + \epsilon_{ij}$$

donde:

y_{ij} = Puntuación obtenida en la variable dependiente por el sujeto i bajo el j -ésimo nivel de la variable de tratamiento.

μ = Media común a todas las observaciones.

η_i = Componente específico asociado al sujeto i y constante a lo largo de las observaciones.

α_j = Efecto debido a la administración del j -ésimo nivel de la variable de tratamiento.

$(\eta\alpha)_{ij}$ = Efecto debido a la interacción entre el i -ésimo sujeto y el j -ésimo nivel de la variable de tratamiento.

ϵ_{ij} = Componente de error específico asociado al sujeto i y al j -ésimo nivel de la variable de tratamiento.

Se asume de ϵ_{ij} es independiente de η_i y que los sujetos han sido seleccionados de una población donde el componente η_i (factor aleatorio) son independientes e idénticamente distribuidas normalmente con media 0 y varianza σ^2_{η} ; los ϵ_{ij} también son independientes e idénticamente distribuidos normalmente con media 0 y varianza σ^2_{ϵ} , y que los niveles del factor de tratamiento son fijos.

Muchas situaciones experimentales son conducidas de manera que a una misma unidad experimental se le aplican sucesivamente varios tratamientos; de donde

resultan valores repetidos de una respuesta sobre la misma unidad u objeto. Los tratamientos pueden ser dietas, dosis de un fármaco, diferentes estímulos, entre otros. Por ejemplo:

- A un animal se le aplican varios medicamentos en diferentes ocasiones o tiempos, luego se le registra su tiempo de pastoreo.
- En pacientes, la tensión arterial sistólica es medida en intervalos de tiempo fijos, como respuesta a un fármaco desde la administración del mismo hasta que aquélla se estabilice.
- Pruebas sobre lectura son administradas a estudiantes en diferentes estadios de su educación, se registran los respectivos puntajes.
- Medidas tales como la alzada y el peso es registrado sobre un tipo de bovino en diferentes edades.
- Medidas sobre la composición del suelo se toman a diferentes profundidades, sobre un terreno experimental.

La información anterior se puede disponer en una matriz $X = (x_{ij})$, donde x_{ij} representa la respuesta a la j -ésima medición (tratamiento) sobre la i -ésima unidad. Las observaciones por fila de esta matriz pueden estar correlacionadas por corresponder a mediciones hechas sobre un mismo sujeto. Si los tratamientos son tales que el orden (temporal o espacial) de aplicación sobre los sujetos puede variarse, entonces la asignación debe aleatorizarse para evitar problemas de sesgo.

Usualmente los individuos pertenecen a grupos distintos o reciben tratamientos diferentes, de manera que uno de los propósitos es estimar o determinar el efecto de los tratamientos sobre las respuestas.

Asumiendo que cada fila, de la matriz $X = (x_{ij})$, es independientemente distribuida respecto a las otras de acuerdo con una normal p -variante con vector de medias

$$\mu' = (\mu_1, \dots, \mu_p),$$

y matriz de covarianzas Σ , se verifica la hipótesis de igualdad de efectos debido a los p -tratamientos; es decir,

$$H_0 : \mu_1 = \dots = \mu_p \text{ frente a: } H_1 : \mu_i \neq \mu_j$$

para algún par $i \neq j = 1, \dots, p$.

Una expresión equivalente a la hipótesis anterior es:

$$H_0 : \begin{pmatrix} \mu_1 - \mu_2 \\ \vdots \\ \mu_{p-1} - \mu_p \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \text{ frente a } H_1 : \begin{pmatrix} \mu_1 - \mu_2 \\ \vdots \\ \mu_{p-1} - \mu_p \end{pmatrix} \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

4.3 ASPECTOS DE DISEÑO

Una de las ventajas del diseño de medidas repetidas es que cada sujeto actúa como su propio control, con lo que se reducen al mínimo los efectos por diferencias individuales. Por ello, el diseño de medidas repetidas es mucho más eficiente. Sin embargo, tal eficiencia no carece de problemas.

4.4 EL ENFOQUE UNIVARIADO

El análisis de varianza con diseño de medidas repetidas de un solo factor se basa en cuatro supuestos:

4.4.1 Supuestos.

a) Muestreo aleatorio independiente. Al igual que en todos los análisis estadísticos, se supone que los sujetos se obtuvieron al azar, o que por lo menos se pueden considerar una muestra aleatoria de la población de interés. Un muestreo aleatorio implica que cualquier observación de la muestra es independiente de cualquier otra, y ésta es una condición previa necesaria para determinar la distribución muestral subyacente. Cabe señalar que este requisito de independencia no implica que las observaciones para una condición de tratamiento no tengan relación con las de otra condición. De hecho, una de las ventajas del diseño de medidas repetidas es que las observaciones por lo general exhiben una correlación positiva, y ello da pie a una estimación de la varianza debida al error más pequeña que la que se obtendría si las observaciones no estuvieran correlacionadas, o tuvieran una correlación negativa. Lo importante es que las diversas repeticiones (es decir, sujetos) sean independientes.

b) Normalidad. Se debe suponer que las distribuciones en las subpoblaciones son normales. Sin embargo, no puede suponerse que las observaciones en una condición de tratamiento sean independientes de las de otro, y esto tiene implicaciones para el análisis.

c) Circularidad de la matriz de varianza / covarianza. El supuesto de circularidad de la matriz de varianza/covarianza implica que la varianza de la

diferencia entre dos subpoblaciones de tratamiento cualesquiera es igual a la varianza de la diferencia entre cualesquiera otras dos.

d) Hipótesis nula. La hipótesis nula para el diseño de medidas repetidas de un solo factor es la misma que para el diseño totalmente aleatorizado de un solo factor, es decir:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$$

Lo anterior establece que las medias de las subpoblaciones de tratamiento son iguales, y por tanto iguales a su media.

4.5 EL ENFOQUE MULTIVARIADO

En el enfoque multivariado, los datos se tratan como un diseño de un solo grupo con variables dependientes, dado que todas las medidas b se miden en la misma escala (es decir, la medida es la misma, simplemente se efectúa en momentos diferentes y en condiciones distintas), se podrían formar nuevas variables consistentes en diferencias entre los pares de variables.

4.5.1 Supuestos.

Esta prueba se basa en tres supuestos:

a) Muestras aleatorias independientes. Al igual que con el enfoque univariado, se supone que las observaciones se obtienen al azar de la población de interés. Esto implica simplemente que las diversas repeticiones son independientes.

b) Normalidad multivariada. Este supuesto implica que no sólo deben estar normalmente distribuidas las variables individuales, sino que sus distribuciones conjuntas también deben ser normales. Aunque éste es un supuesto necesario, se ha demostrado que el análisis de varianza multivariado es robusto respecto a violaciones del supuesto.

c) Hipótesis nula. La hipótesis nula podría plantearse en términos multivariados así:

$$H_0: \mu_1 - \mu_4 = \mu_2 - \mu_4 = \mu_3 - \mu_4 = 0$$

es decir, la diferencia entre las medias de la población 1 y 4 es igual a la diferencia entre las poblaciones 2 y 4 e igual a la diferencia entre las poblaciones 3 y 4, y es igual a cero. La matriz se puede expresar así:

$$\begin{vmatrix} \mu_1 - \mu_4 \\ \mu_2 - \mu_4 \\ \mu_3 - \mu_4 \end{vmatrix} = 0$$

Un factor intra-sujetos o de medidas repetidas se caracteriza porque todos los niveles del factor se aplican a los mismos sujetos.

Entre las ventajas de los diseños de medidas repetidas están:

- Se requieren menos sujetos que en un diseño completamente aleatorizado.
- Permiten eliminar la variación residual debida a las diferencias entre los sujetos, pues se utilizan los mismos durante el experimento.

Como contrapartida, es necesario vigilar algunos efectos atribuibles precisamente a la utilización de los mismos sujetos, tales como el efecto de arrastre, que ocurre cuando se administra una condición antes de que haya finalizado el efecto de otra administrada previamente; o el efecto del aprendizaje por la práctica, que ocurre cuando las respuestas de los sujetos pueden mejorar con la repetición y, como consecuencia de ello, los tratamientos administrados en último lugar parecen más efectivos que los administrados en primer lugar, sin que haya diferencias reales entre ellos. Obviamente, conviene conocer las ventajas e inconvenientes de estos diseños para decidir correctamente cuándo es apropiado utilizarlos.

Para el modelo a elaborar se evaluará la interacción entre el canal y el tiempo. Se determinará si el tiempo es un factor determinante en las ventas en un cliente (j), en un canal (i) y un tiempo (k).

5 ANÁLISIS DE DATOS

Para efectos del trabajo, el análisis se concentrará en el comportamiento de la Regional Centro (Bogotá y poblaciones aledañas) y sobre esta misma se elaborará el modelo de medidas repetidas.

Haciendo un análisis descriptivo general de la Empresa de los datos mensuales (enero a diciembre) de los años 2013 y 2014 se encuentra que:

Tabla 1. Análisis Descriptivo General de Empresa (ventas en MM)

Medida	Parámetro
Media	7.935,38
Error típico	221,87
Mediana	7.816,79
Moda	0
Desviación estándar	1.086,94
Varianza de la muestra	1.181.447,79
Curtosis	-0,092552093
Coefficiente de asimetría	0,174567012
Rango	4.406,81
Mínimo	5.710,42
Máximo	10.117,23
Suma	190.449,19
Cuenta	24

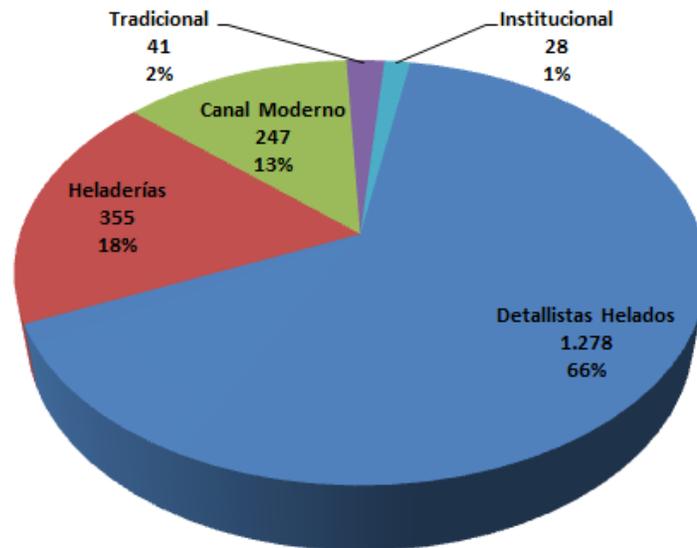
Con el modelo de medidas repetidas se podrá ratificar la importancia del tiempo en la evolución de las ventas, volviéndose éste un factor determinante al momento de pronosticar y hacer proyecciones de éstas.

Se podría suponer que en canal como el Institucional en donde se abarcan clientes de temporadas, el tiempo deberá ser determinante en el momento de pronosticar, y seguramente el modelo de medidas repetidas dejará ver la relación entre este canal, las ventas y el tiempo.

En un canal como el TAT que está conformado por establecimientos comerciales con alto tráfico de clientes como cadenas de droguerías, colegios, universidades, panaderías, entre otros, se esperaría que las ventas marquen una tendencia en el tiempo, dado el comportamiento de los individuos que rodean este tipo de negocios.

5.1 ANÁLISIS DESCRIPTIVO DE LA MUESTRA

Calculando la participación de los canales en los datos de estudio se encontró que entre el canal Detallistas y Heladerías hay una concentración del 83,79% de los clientes.



Gráfica 5. Análisis Descriptivo de la Muestra (Participación de Clientes por Canales)

Los datos muestran que la empresa hace un gran esfuerzo por conquistar el Canal Detallistas abarcando este segmento de mercado para realización de sus ventas. De la misma manera, se observa que el Canal Heladerías tiene gran potencial para su crecimiento; las heladerías son una buena fuente de trabajo, y su naturaleza es la venta exclusiva de helado y de productos preparados a base de este insumo.

Por el contrario, en el Canal Moderno se podría esperar una evolución más lenta, ya que los negocios en los cuales se posiciona este producto no tiene como naturaleza exclusiva la venta del mismo, encontrando el consumidor otros productos, posiblemente sustitutos, con los cuales satisfacer su necesidad, puesto que estos compradores, están más conscientes de la compra que quieren realizar, ya que este tipo de negocios no se enfoca en productos de impulso sin en productos que son para el hogar.

En cuanto al Canal Tradicional, se observa que la participación es pequeña. Esto es debido a que este tipo de negocios corresponden a distribuidores que no venden directamente al consumidor final sino que llegan a detallistas de poblaciones aledañas. Por lo tanto, para ser un distribuidor se necesita contar con

un capital que permita tener una estructura logística para llegar a estos puntos de venta.

5.2 MUESTREO

Teniendo en cuenta que se tienen diferentes tipos de canales de distribución que operan de diversas maneras y con diferentes tipos de clientes se optó por trabajar con un muestreo EST-MAS (muestreo aleatorio simple dentro de cada estrato).

El procedimiento de estimación más frecuente en el muestreo estratificado consiste en considerar cada estrato como una subpoblación y realizar, en primera instancia, estimaciones acerca de los parámetros correspondientes a cada una de esas subpoblaciones. Una vez se han hecho las estimaciones para estos subgrupos, se procede a combinarlas para obtener las estimaciones globales de los parámetros de interés. La estimación dentro de cada estrato puede llevarse a cabo de acuerdo con diferentes procedimientos, lo importante es que las muestras seleccionadas en cada uno de los estratos sean independientes para poder obtener fórmulas directas de estimación para los parámetros poblacionales.³

5.2.1 Requisitos de una buena muestra.

Una buena muestra reproduce las características de interés en la población, lo más cerca posible. Será representante en el sentido de que cada unidad muestreada representa las características de un número conocido de unidades en la población.

En una encuesta ideal, la población de la muestra será idéntica a la población objetivo, pero este ideal rara vez se cumple exactamente. En las encuestas de las personas, la población muestreada es generalmente menor que la población objetivo.

Existen diferentes tipos de muestras de probabilidad. Los términos “muestra aleatoria simple”, “muestra estratificada” y “muestra por conglomerados” son básicas para cualquier discusión de encuestas por muestreo.

a. Una muestra aleatoria simple (MAS) es la forma más simple de la muestra probabilística. Un MAS de tamaño n es tomada cuando cada subgrupo posible de n unidades de la población tiene la misma probabilidad de ser la muestra. En la toma de una muestra aleatoria, el investigador está en efecto mezclando la

³ LOHR, Sharon L. Sampling: Design and Analysis. Segunda Edición. Arizona State University . p. 2, 23-35, 95-103.

población antes de agarrar n unidades. El investigador no necesita examinar cada miembro de la población por la misma razón que un técnico médico no necesita drenar de sangre para medir el conteo de glóbulos rojos: su sangre se mezcla suficientemente bien que cualquier muestra debe representante.

b. En un muestreo simple estratificado (EST-MAS), la población se divide en subgrupos llamados estratos. Un MAS es seleccionado de cada estrato, y el MAS en los estratos se seleccionan independientemente. Los estratos suelen ser subgrupos de interés para el investigador.

c. En una muestra de conglomerados, las unidades de observación de la población se agrupan en unidades de muestreo más grandes, llamados racimos.

Todos estos métodos – MAS, muestreo aleatorio estratificado y muestreo de conglomerados – implica la selección aleatoria de unidades en la muestra. En un MAS, las unidades de observación se seleccionan al azar de la población de unidades de observación; en una muestra aleatoria estratificada, unidades de observación en cada estrato se seleccionan aleatoriamente; en una muestra de clúster (conglomerado), el clúster se seleccionan aleatoriamente de la población de todos los clústeres. Cada método es una forma de muestreo probabilístico.

5.2.2 Muestreo Aleatorio Simple.

El MAS es la forma más básica del muestreo probabilístico y proporciona la base teórica de las formas más complicadas. Hay dos maneras de tomar una muestra aleatoria simple: con reemplazo, en el cual la misma unidad puede incluir más de una vez en la muestra y sin el reemplazo, en el que todas las unidades de la muestra son diferentes.

Una muestra aleatoria simple con reemplazo de tamaño n de una población de N unidades puede considerarse un dibujo de n muestras independientes de tamaño 1. Una unidad se selecciona aleatoriamente de la población para ser la primera unidad muestreada, con probabilidad $1/N$. Entonces se sustituye la unidad muestreada en la población, y una segunda unidad es seleccionada al azar con probabilidad $1/N$. Este procedimiento se repite hasta que la muestra tenga n unidades, la cual puede incluir duplicados de la población.

En el muestreo de población finita, sin embargo, se muestrea dos veces la misma persona lo cual no brinda ninguna información adicional. Generalmente se prefiere el método sin reemplazo de la muestra para que la muestra no contenga duplicados. Una muestra aleatoria simple sin reemplazo de tamaño n se selecciona para que cada subgrupo posible de n unidades distintas de la población

tenga la misma probabilidad de ser seleccionadas como muestra. Hay $\binom{N}{n}$ muestras posibles y cada una es igual de probable, por lo que la probabilidad de seleccionar cualquier muestra individual S de n unidades es

$$P(S) = \frac{1}{\binom{N}{n}} = \frac{n!(N-n)!}{N!}.$$

Como consecuencia de esta definición, la probabilidad de que cualquier unidad dada aparece en la muestra es n / N .

Para tomar un MAS, se necesita una lista de todas las unidades de observación en la población; esta lista es el marco de muestreo. En un MAS, la unidad de muestreo y observación coinciden. Cada unidad tiene asignado a un número, y se seleccionó una muestra para que (1) cada unidad tenga la misma probabilidad de ocurrir en la muestra y (2) la selección de una unidad no sea influenciada por otras unidades que ya han sido seleccionadas. En la práctica, los números pseudoaleatorios generados por computadora se utilizan generalmente para seleccionar una muestra.

5.2.3 Muestreo Estratificado.

Se divide la población de unidades de muestreo N en H "capas" o estratos, con N_h unidades de muestreo en el estrato h . Para trabajar con el muestreo estratificado, debemos conocer los valores de N_1, N_2, \dots, N_H y debe tener

$$N_1 + N_2 + \dots + N_H = N,$$

Donde N es el número total de unidades en toda la población.

En el muestreo aleatorio estratificado, la forma más simple de muestreo estratificado, se toma independientemente como un MAS de cada estrato para que n_h observaciones sean seleccionadas al azar de las unidades de población en el estrato h . Se define S_h para ser el conjunto de unidades de n_h en el MAS para estrato h .

Notación para la estratificación: Las cantidades de población son:

y_{hj} = valor de la unidad j en el estrato h

$t_h = \sum_{j=1}^{N_h} y_{hj}$ = población total en el estrato h

$t = \sum_{h=1}^H t_h$ = Población total

$$\bar{y}_{hU} = \frac{\sum_{j=1}^{N_h} y_{hj}}{N_h} = \text{Media poblacional en el estrato } h$$

$$\bar{y}_U = \frac{t}{N} = \frac{\sum_{h=1}^H \sum_{j=1}^{N_h} y_{hj}}{N} = \text{Media poblacional general}$$

$$S_h^2 = \sum_{j=1}^{N_h} \frac{(y_{hj} - \bar{y}_{hU})^2}{N_h - 1} = \text{Varianza de la población en el estrato } h$$

Cantidades correspondientes para la muestra, usando estimaciones MAS dentro de cada estrato, son:

$$\bar{y}_h = \frac{\sum_{j \in \mathcal{S}_h} y_{hj}}{n_h}$$

$$\hat{t}_h = \frac{N_h}{n_h} \sum_{j \in \mathcal{S}_h} y_{hj} = N_h \bar{y}_h$$

$$s_h^2 = \sum_{j \in \mathcal{S}_h} \frac{(y_{hj} - \bar{y}_h)^2}{n_h - 1}$$

Suponga que sólo muestra el estrato de h . En efecto, tenemos una población de N_h unidades y se toma un MAS de n_h unidades. Luego se calcularía \bar{y}_{hU} para \bar{y}_h , y para $\hat{t}_h = N_h \bar{y}_h$. El total de población es de $t = \sum_{h=1}^H t_h$, por lo que se calcula t para

$$\hat{t}_{\text{str}} = \sum_{h=1}^H \hat{t}_h = \sum_{h=1}^H N_h \bar{y}_h.$$

Para calcular entonces \bar{y}_U , se usa

$$\bar{y}_{\text{str}} = \frac{\hat{t}_{\text{str}}}{N} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h.$$

Esto es un promedio ponderado de los promedios de estrato de muestra; los pesos son los tamaños relativos de los estratos. Para usar el muestreo estratificado, deben conocerse el tamaño o tamaños de familiares de los estratos.

Las propiedades de los estimadores siguen directamente las propiedades de los estimadores del MAS:

- Insesgamiento. \bar{y}_{str} and \hat{t}_{str} son estimadores insesgados de \bar{y}_U and t_U . Esto es cierto porque

$$E \left[\sum_{h=1}^H \frac{N_h}{N} \bar{y}_h \right] = \sum_{h=1}^H \frac{N_h}{N} E[\bar{y}_h] = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_{hU} = \bar{y}_U.$$

- La varianza de los estimadores. Como se tiene un muestreo independientemente de los estratos y se sabe $V(\hat{t}_h)$ de la teoría del MAS, las propiedades del valor esperado y la ecuación implica que

$$V(\hat{t}_{str}) = \sum_{h=1}^H V(\hat{t}_h) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{S_h^2}{n_h}.$$

- Estimaciones de varianza para muestras estratificadas. Se puede obtener un estimador insesgado de $V(\hat{t}_{str})$ mediante la sustitución de la muestra estimada s_h^2 para las cantidades de población S_h^2 . Se nota que, para estimar las varianzas, se necesitan al menos dos unidades de cada estrato de la muestra:

$$\hat{V}(\hat{t}_{str}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h},$$

$$\hat{V}(\bar{y}_{str}) = \frac{1}{N^2} \hat{V}(\hat{t}_{str}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h}.$$

Como siempre, el error estándar de un estimador es la raíz cuadrada de la varianza estimada:

$$SE(\bar{y}_{str}) = \sqrt{\hat{V}(\bar{y}_{str})}.$$

- Intervalos de confianza para muestras estratificadas. Si (1) los tamaños de muestra en cada estrato son grandes o (2) el diseño de muestreo tiene una larga serie de estratos, un aproximado intervalo de confianza (IC) $100(1 - \alpha)\%$ para la media es

$$\bar{y}_{str} \pm z_{\alpha/2} SE(\bar{y}_{str}).$$

Para el caso de estudio se la aplica la fórmula general de cálculo de la muestra, haciendo una estimación inicial de la misma:

$$n_0 = Z^2 \alpha^2 / \varepsilon^2$$

Para la estimación de la muestra inicial se considerarán los siguientes parámetros:

$$\varepsilon = 5\%$$

$$Z = 1,96$$

$$\alpha^2 = 0,25$$

Reemplazando los anteriores valores en la fórmula el resultado sería:

$$n_0 = (1,96)^2 * 0,25^2 / 0,05^2$$

$$\approx 384,16$$

Una vez obtenido el valor inicial de la muestra de la fórmula general, se procedió a corregir el tamaño de la muestra teniendo en cuenta que se conoce el tamaño de la población.

Se elaborará un EST-MAS para determinar el tamaño de la muestra y hacer la distribución de los clientes que se tomarán en cada canal.

Para efectos de la determinación de la muestra se tomaron los clientes que tuvieron continuidad de la venta durante 24 meses continuos. Partiendo de ellos se calculó la cantidad de clientes de los diferentes canales de ventas, y con ellos se calculó la proporción (participación) de los clientes por canal.

Tabla 2. Muestreo Aleatorio Simple Estratificado (EST-MAS)

Canal de Distribución	Muestra
Detallistas Helados	210
Heladerías	58
Canal Moderno	41
Tradicional	7
Institucional	5
Muestra Total (<i>n</i>)	321

De acuerdo a los datos obtenidos se necesitarán 321 registros para la elaboración del modelo de Medidas Repetidas. De la muestra obtenida el estrato con mayor muestra será el canal de distribución Detallistas, con lo que se concluye que es ahí donde hay mayor número de clientes contribuyendo a la venta.

5.3 ANÁLISIS POR CANAL DE DISTRIBUCIÓN

Se observa que el canal Tradicional es el canal con promedio más alto y esto puede deberse a que tiene pocos clientes que manejan ventas de alto volumen.

Tabla 3. Promedio por Canal de Distribución

		Venta
		Media
Canal de distribución	Canal Moderno	978,079
	Detallistas Helados	231,817
	Heladerías	694,596
	Institucional	472,832
	Tradicional	8.208,363

Considerando el valor de la Desviación Típica se puede decir que los datos no están muy dispersos y que por lo tanto no están concentrados alrededor de la media muestral. Este resultado puede ser consistente dada la variabilidad de los canales de distribución.

Tabla 4. Estadísticos por Canal de Distribución

	Canal de distribución	Venta
N Válidos	7704	7704
Perdidos	0	0
Media		588,44881
Mediana		265,04050
Desv. típ.		2560,424264
Suma		4533409,650

En cuanto al grupo de clientes se observa que las tiendas son el grupo con mayor cantidad de clientes y que tiene una participación del 38%, seguidas de las

Heladerías con un 11%. Se determina que el 40% del grupo de clientes están participando con el 80% de los clientes de la muestra.

Tabla 5. Estadísticos por Grupo de Clientes

	Frecuencia	Porcentaje	Porcentaje acumulado
Tiendas	2928	38	38,0
Heladerias	840	10,9	48,9
Kam Éxito	384	5	53,9
Supermercados	384	5	58,9
Indepe	288	3,7	62,6
Cigarrerias	288	3,7	66,3
Heladeria Premium	264	3,4	69,7
Copidrogas	264	3,4	73,1
Fruterias	240	3,1	76,2
Restaurantes	216	2,8	79,0
Cabinas/Salas Intern	216	2,8	81,8
Droguería Droxi	216	2,8	84,6
Droguerias	216	2,8	87,4
Informal	192	2,5	89,9
Panaderias	144	1,9	91,8
Minimercados	120	1,6	93,4
Kam Olímpica	120	1,6	95,0
Microaliado	96	1,2	96,2
Heladero	72	0,9	97,1
Cafeterias	72	0,9	98,0
Comercio Directo	48	0,6	98,6
Resto Cadenas	24	0,3	98,9
Distribuidor Institu	24	0,3	99,2
Aliados Con Cava	24	0,3	99,5
Cadenas Droguerias	24	0,3	100
Instituciones Educat	24	0,3	
Mayoristas	24	0,3	
Heladeros	24	0,3	
Total	7704	100	

5.4 MODELO

Considerando las variables *ventas*, *grupo de clientes*, *canal de distribución*, su comportamiento en el tiempo y con la ayuda de la herramienta RStudio se plantea la siguiente hipótesis considerando:

H_0 : El tiempo no se rechaza como factor determinante en las ventas de consumo masivo, durante el periodo comprendido entre los años 2012 y 2014

$$(H_0 : k_1=0, k_2=0, k_3=0, k_4=0, k_5=0 \dots k_{24}=0, = 0)$$

H_a : El tiempo no es como factor determinante en las ventas de consumo masivo, durante el periodo comprendido entre los años 2012 y 2014.

$$(H_a : \text{Existe } i \neq j \text{ t } k \text{ } k_i \neq k_j)$$

Considerando las variables de estudio y el error como uno de los términos de la ecuación lineal, se obtiene:

$$y_{ijk} = \mu + \alpha_i + \kappa_k + (\alpha\kappa)_{ik} + \varepsilon_{ijk}$$

α_i = Efecto del i ésimo canal

κ_k = Efecto del k ésimo tiempo

$(\alpha\kappa)_{ik}$ = Interacción del canal con el tiempo

j = Cliente

Al reemplazar en la ecuación anterior los datos de las variables y de acuerdo a los resultados obtenidos se concluye que se acepta H_0 , por lo cual el tiempo es un factor determinante en el comportamiento de las ventas del periodo en estudio.

```

>
> ##Modelo
> Sujeto=as.factor(Cliente)
> Tiempo=as.factor(Tiempo)
> fit=aov(Venta~Canal.de.distribución*Tiempo+Error(Sujeto))
> summary(fit)

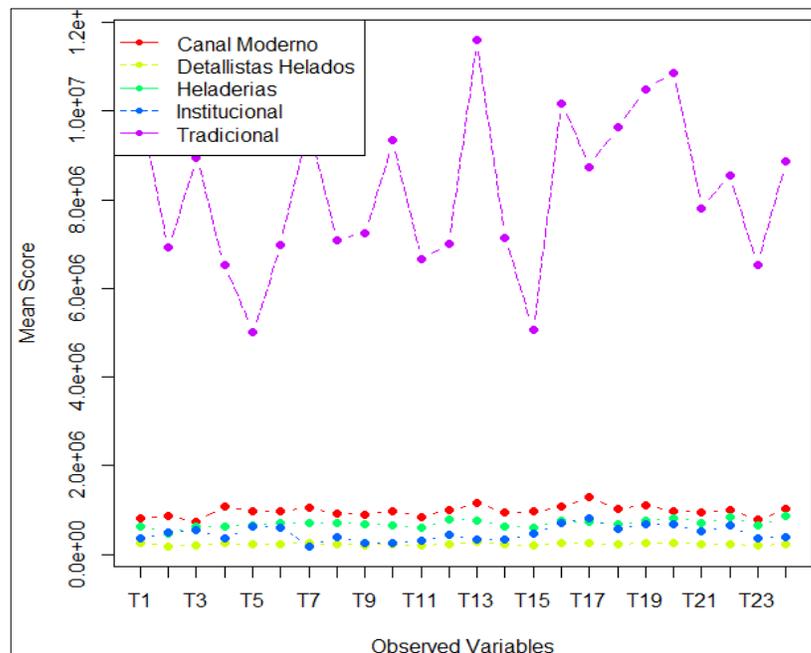
Error: Sujeto
              Df Sum Sq Mean Sq F value Pr(>F)
Canal.de.distribución  4 1.056e+16 2.641e+15  21.38 1.66e-15 ***
Residuals              296 3.656e+16 1.235e+14
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: within
              Df Sum Sq Mean Sq F value Pr(>F)
Tiempo              23 3.69e+13 1.604e+12  4.132 1.21e-10 ***
Canal.de.distribución:Tiempo  92 5.13e+14 5.576e+12  14.361 < 2e-16 ***
Residuals          7288 2.83e+15 3.883e+11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Como el *P*valor (4.132) es mayor al nivel de significancia (0.05) se concluye que se acepta que el tiempo es un factor determinante en las ventas de productos de consumo masivo, durante el periodo comprendido entre los años 2013 y 2014.

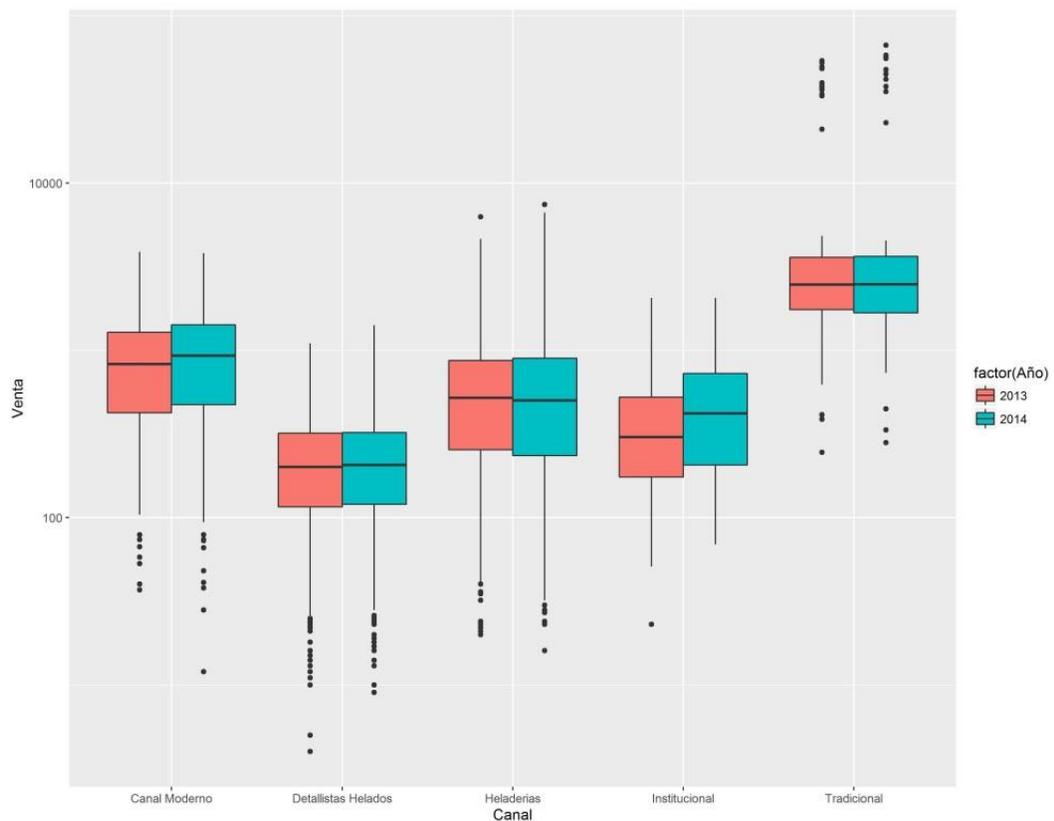
$P_v > \alpha$, se acepta H_0



Gráfica 6. Comportamiento de los canales a través del tiempo

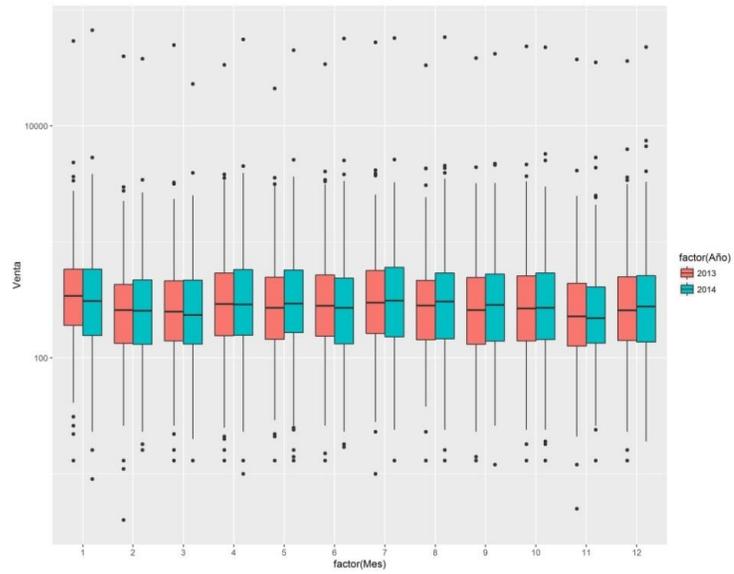
En los siguientes gráficos se hizo una transformación logarítmica en las ventas para poder evaluar el comportamiento de las mismas sabiendo que hay valores muy altos en las ventas, por lo tanto, se precisa dicha transformación logarítmica para poder apreciar el comportamiento de los datos.

Se observa que las ventas a través del tiempo se comportan de manera similar, tanto por mes como por canal. Sin embargo en el Canal Institucional se observan algunos cambios, dado que este canal está sujeto a adjudicación de licitaciones y negociaciones especiales que pueden o no darse durante el año.



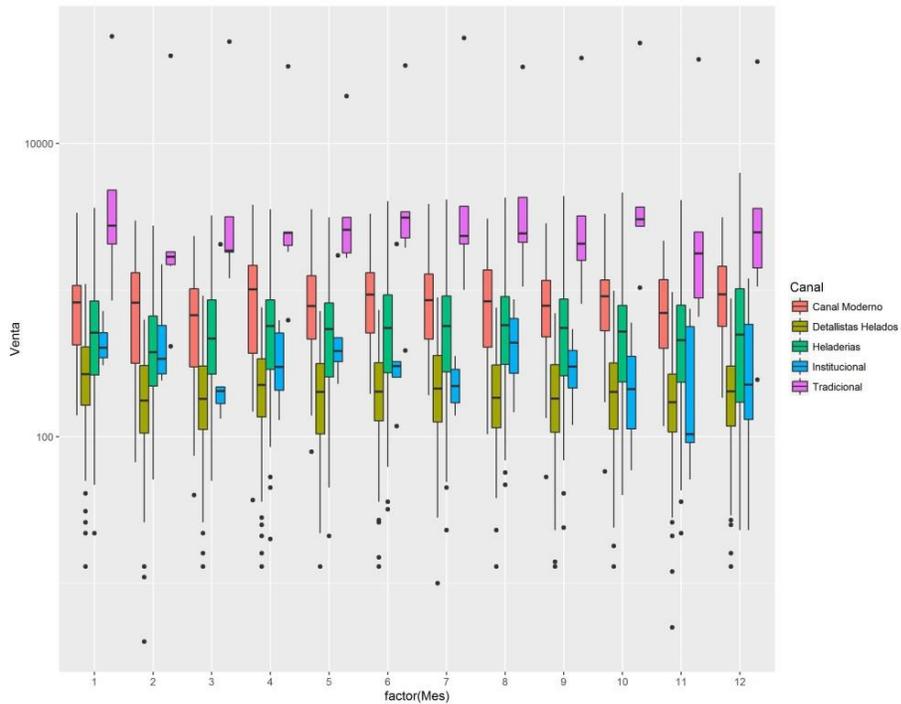
Gráfica 7. Gráfico de cajas y bigotes de las Ventas Vs Canal de Distribución

En la Gráfica 7 se observa que los datos agrupados por año mantienen su comportamiento entre uno y otro periodo; no se detectan grandes variabilidades que permitan extraer cambios concluyentes y diferenciadores con respecto al comportamiento visto en las gráficas iniciales. Sin embargo, en Canal Institucional deja ver un desplazamiento en la concentración de los datos de un año a otro, esto dato a que en este canal las ventas se hacen de acuerdo a adjudicaciones anuales, convenios y algunos negocios puntuales que posiblemente no se repiten año tras año.

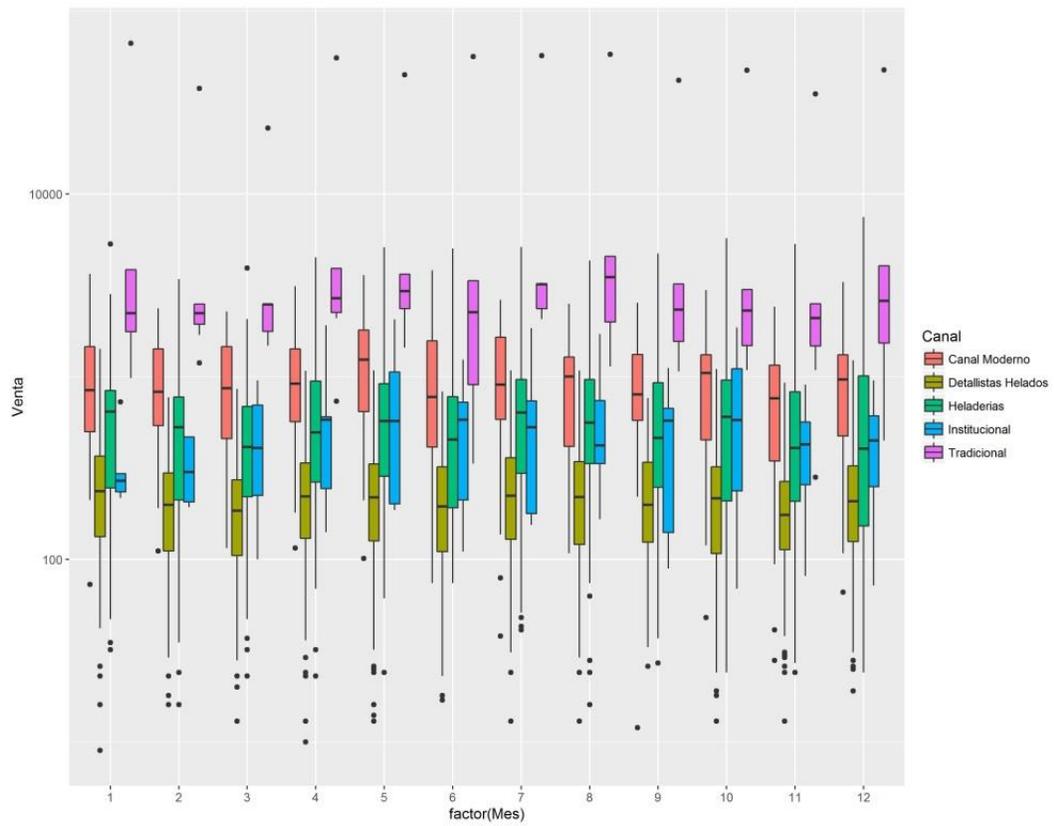


Gráfica 8. Gráfico de cajas y bigotes de las Ventas Vs Periodos de Venta

En la Gráfica 8 se puede observar que cuando todos los datos se miran en su conjunto a través de los periodos a analizar, es decir, que no hay distinción por canal el comportamiento de los datos tanto en el año 2013 como con el 2014 se mantiene homogéneo.



Gráfica 9. Gráfico de cajas y bigotes de las Ventas por Canal Vs Periodos de Venta año 2013



Gráfica 10. Gráfico de cajas y bigotes de las Ventas por Canal Vs Periodos de Venta año 2014

En las Gráficas 9 y 10 se observa que a pesar de que los datos por canal se mantienen homogéneos con las gráficas iniciales, la distribución de los datos cambia de manera notoria en los Canales Tradicional e Institucional encontrando una variabilidad intercuartílica importante.

6 CONCLUSIONES

Este trabajo permitió entender la importancia de la estadística en el análisis de las cifras, rescatando la aplicación de modelos estadísticos que permitan ahondar en las variables que podrían ser parte de un modelo de predicción.

Para diseñar un modelo de medidas repetidas se hace necesario aplicar métodos de muestreo que permita contar con datos óptimos y suficientes con los cuales llegar al mejor resultado.

Los datos trabajados permitieron estimar valores estadísticamente significativos con los cuales se pudieron concluir que efectivamente el tiempo es un factor determinante a la hora de establecer modelos de predicción en la Empresa, ya que el mismo sí influye en el comportamiento de las ventas.

Al comparar las gráficas iniciales versus las gráficas de cajas se notan cambios en estas últimas en el comportamiento entre los periodos de un año a otro, a diferencia del comportamiento homogéneo que se observa en las primeras.

El modelo de medidas repetidas es aplicado principalmente en experimentos del área de la salud y farmacología. No son modelos aplicados comúnmente para analizar el comportamiento de las ventas, razón por la cual no se consigue fácilmente teoría aplicada en este campo.

Como análisis en el comportamiento de las ventas se hace suposición de que cada variable influirá en las ventas de la misma manera y que las ventas variarán en función del periodo y en función de cada canal.

BIBLIOGRAFÍA

- BALLUERKA, N. & VERGARA A.I. (2002). *Diseños de Investigación Experimental en Psicología*. España: Pearson Educación S.A.
- LOHR, S.L. (1999). *Sampling: Design and Analysis* (2da Ed.). EE.UU: Arizona State University.
- OSPINA, D. (2001). *Introducción al Muestreo* (1ra Ed.). Bogotá, Colombia: Unibiblos.
- PASCUAL, J., FRÍAS, D. & GARCÍA, F. (1996). *Manual de Psicología Experimental: metodología de investigación* (1ra Ed.). Barcelona, España: Ariel Psicología.
- VÁZQUEZ, D. (2014). Mercados de Consumo. *Mkt Diego Vázquez Magazine*. Recuperado el 13 de agosto de 2015, en <http://www.mktdiegovazquez.com/>