

Material de consulta

Estructurado en siete temas
Este corresponde al Tema 4

Elaborado por:
Econ. Milton Oroche Carbajal

**TEMA 1:
INVESTIGACIÓN DE MERCADOS.**

**TEMA 2:
MÉTODOS CUALITATIVOS DE INVESTIGACIÓN DE MERCADOS.**

**TEMA 3:
MÉTODOS CUANTITATIVOS DE INVESTIGACIÓN DE MERCADOS.**

**TEMA 4:
MÉTODOS DE ANÁLISIS DE LA INFORMACIÓN: ANÁLISIS
UNIVARIABLE Y BIVARIABLE.**

- 4.1 Introducción. Codificación y Tabulación de Datos en el Análisis de la Información
- 4.2. Métodos de Análisis de la Información Obtenida en la Investigación Comercial
- 4.3. Métodos de Análisis Univariantes y Bivariantes
- 4.4. Métodos de Inferencia Estadística.

**TEMA 5:
ANÁLISIS CAUSAL: LA EXPERIMENTACIÓN COMERCIAL.**

**TEMA 6:
EL ANÁLISIS MULTIVARIABLE Y SUS APLICACIONES AL
MARKETING.**

**TEMA 7:
DISEÑO Y PRESENTACIÓN DE INFORMES DE INVESTIGACIÓN.**



TEMA 4: MÉTODOS DE ANÁLISIS DE LA INFORMACIÓN: ANÁLISIS UNIVARIABLE Y BIVARIABLE.

4.1. INTRODUCCIÓN. CODIFICACIÓN Y TABULACIÓN DE DATOS EN EL ANÁLISIS DE LA INFORMACIÓN.

➤ **INTRODUCCIÓN**

En un proceso de investigación de mercados una vez que la información está recogida, el paso siguiente es adoptar una metodología concreta o específica de análisis de datos con el fin de garantizar que las decisiones que se toman, se realicen con una base que sea racional.

Adoptar una metodología de análisis de datos supone trabajar o apoyarse en la estadística. Dentro de la estadística se pueden distinguir dos partes bien diferenciadas: la estadística descriptiva y la inferencia estadística. La estadística descriptiva tiene como objetivo sintetizar la información contenida en los elementos de una muestra.

Por lo tanto, la estadística descriptiva sólo nos proporciona lo que se denomina medidas resumen como por ejemplo: la media, la moda, la mediana, la desviación típica siendo estas medidas de posición y dispersión. Es muy importante dejar claro que la información sólo es válida para la muestra y no para toda la población.

El objetivo de la inferencia estadística es generalizar o inferir los datos de la muestra al conjunto de la población investigada o estudiada. Para poder aplicar la técnica de la inferencia estadística es imprescindible contar con una muestra que sea representativa de la población objeto de estudio.

➤ **CODIFICACIÓN Y TABULACIÓN DE DATOS EN EL ANÁLISIS DE LA INFORMACIÓN**

Para poder aplicar los métodos de la estadística descriptiva o de la inferencia estadística, lo primero que hay que hacer es asignar valores concretos a las opiniones o comportamientos de las personas, empresas, marcas que se estén estudiando, dejando bien claro que asignar valores no es más que realizar mediciones. A continuación, se van a ver distintos tipos de escalas o medidas, concretamente cuatro, que se pueden utilizar para asignar valores:

1- Escalas nominales:

Este tipo de escala consiste en clasificar en categorías a las personas, empresas, marcas, es decir, a los elementos que se estén estudiando o que son objeto de medida. Posteriormente, se asigna un número a cada una de las categorías y se considera que todos los elementos a los que se ha asignado el mismo número son cualitativamente idénticos en la variable de medida. Con las escalas nominales sólo se puede saber si un elemento es igual o distinto a otros elementos de la muestra. Las categorías tienen que cumplir dos requisitos para que estén bien diseñadas:

- a) Que sean exhaustivas: esto significa que no puede haber ningún elemento que quede sin clasificar en alguna de las categorías que se han fijado.
- b) Que sean exclusivas: lo que significa que sólo se puede estar en una y sólo una de las categorías.

Las escalas nominales son, especialmente, útiles en el caso de las variables cualitativas, funcionando los números asignados únicamente como etiquetas identificativas.

Algunos ejemplos de variables cualitativas a las que se podría aplicar esta escala son las siguientes: profesión, nivel de estudios, sexo, edad, ingresos, comunidad autónoma del que procede el encuestado, estado civil, si una persona conoce o no una determinada marca, etc.

2- Escalas ordinales:

Este tipo de escala consiste en asignar a los elementos medidos un número que permita ordenarlos según la cantidad de variable que poseen desde el punto de vista del encuestado. Las escalas ordinales son útiles principalmente para variables cualitativas y aquí los números permiten afirmar si la cantidad de variable que posee un elemento es mayor o menor que la de otro, pero no dice cuanto mayor o cuanto menor, además de saber si un elemento es igual o distinto a otros elementos de la muestra.

Un ejemplo típico es el de indicar el orden de preferencia de las marcas, de unos productos, etc. Con los números se va a poder decir que una marca es más o menos preferible a otra, pero no cuanto más o menos preferida. Otro ejemplo son los rankings de preferencia.

3- Escalas de intervalo:

Las escalas de intervalo consisten en definir una unidad de medida y después asignar a cada elemento medido un número indicativo de la cantidad de variable que posee según la unidad de medida establecida.

En las escalas de intervalo no existe el cero absoluto de forma que la diferencia entre los elementos no son constantes. Esta escala se utiliza para medir el grado de satisfacción y el grado de acuerdo o desacuerdo con unas determinadas afirmaciones.

Este tipo de escala es adecuada para variables cuantitativas y se utiliza en gran medida para la evaluación de actitudes. En esta escala los números permiten afirmar si un elemento es igual o distinto a otros elementos de la muestra, si la cantidad de variable que posee un elemento es mayor o menor que la de otro, pero en este caso si nos dice cuanto mayor o cuanto menor.

4- Escalas de razón:

Las escalas de razón son como las escalas de intervalos pero añadiéndoles el cero absoluto, es decir, la ausencia total de

cantidad de variable. Este tipo de escalas están indicadas para variables cuantitativas.

En esta escala los números permiten afirmar si un elemento es igual o distinto a otros elementos de la muestra, si la cantidad de variable que posee un elemento es mayor o menor que la de otro, pudiendo decir cuanto mayor o cuanto menor.

La diferencia entre los elementos medidos son constantes y, además, se puede afirmar si la cantidad de uno es el doble, el triple, etc. que la cantidad de otro elemento. Algunos ejemplos de elementos a los que se podría aplicar escalas de razón son: los ingresos, la edad, los volúmenes de venta, etc.

La información sobre una misma variable se puede recoger mediante cualquiera de las cuatro escalas vistas, aunque eso si, unas son más adecuadas para unos casos que para otros.

Escalas concretas para la medición de actitudes:

Nos detenemos en estos tipos de escalas porque la tarea de medición de actitudes es muy compleja ya que es algo que no se puede observar directamente. Las actitudes se suelen medir, generalmente, a través de las escalas de intervalo pero con algunas características particulares.

a) Escala de clasificación por sumas constantes:

Esta escala consiste en asignar un número de puntos a distintos atributos según su importancia. Por ejemplo, 100 puntos a repartir entre cuatro atributos por orden de importancia. El inconveniente es que este tipo de escala es sólo útil cuando se trabaja con un número pequeño de atributos.

b) Escala diferencial-semántico:

Esta escala consiste en utilizar palabras o frases bipolares, es decir, términos o extremos opuestos para que el entrevistado exprese sus sentimientos indicando su posición entre esos extremos.

Ejemplo. Los atributos a considerar son los tres siguientes:
X, Y, Z.

	-3	-2	-1	0	+1	+2	+3	
X: Muy bien decorado						x		Muy mal decorado
Y: Soluciona mis problemas						x		No soluciona mis problemas
Z: Me inspira confianza						x		No me inspira confianza

Este tipo de escala permiten diseñar perfiles y realizar comparaciones entre los distintos elementos de la investigación, además de analizar los puntos débiles y fuertes de cada elemento. La escala que se utiliza va desde el -3 hasta el +3. Por último, hay que decir que este tipo de escala se utiliza mucho para estudiar la imagen de las marcas.

c) Escala de Stapel:

En este tipo de escala se utiliza solo una palabra, adjetivo o frase. Es una variación de la escala diferencial-semántico, ya que ahora las puntuaciones van de -5 a +5, de forma que cuanto más se aproxime el término a la realidad más positiva será la puntuación que reciba.

Ejemplo:

-5	-4	-3	-2	-1	1	2	3	4	5
x (suave)									Sabor Fuerte
x (más barato)									Precio alto
Color claro									x (bastante claro)

Este tipo de escala sirve para lo mismo que la escala diferencial-semántico, es decir, para el diseño de perfiles, la comparación de elementos, etc.

d) Escala Likert:

Este tipo de escala consiste en pedirle al entrevistado que indique su grado de acuerdo o desacuerdo en relación a una serie de afirmaciones en una escala que va de 1 a 5, donde el 1 significa que se está totalmente en desacuerdo con la afirmación, el 2 significa que se está en desacuerdo, el 3 ni en acuerdo ni en desacuerdo, el 4 significa que se está de acuerdo

con la afirmación y el 5 totalmente en acuerdo. En la escala Likert los atributos más interesantes son los que discriminan mucho entre los encuestados.

4.2. MÉTODOS DE ANÁLISIS DE LA INFORMACIÓN OBTENIDA EN LA INVESTIGACIÓN COMERCIAL.

Los métodos de análisis de la información obtenida en la investigación comercial se pueden dividir entre los de la estadística descriptiva y los de la estadística diferencial. Dentro de la estadística descriptiva hay dos grupos de métodos: univariables y bivariables.

Los métodos univariables analizan información para una sola variable mientras que los métodos bivariables analizan información simultáneamente para dos variables.

Dentro de los métodos univariables hay que distinguir entre medidas de posición y medidas de dispersión. Dentro de las medidas de posición se puede distinguir entre: medidas de tendencia central, medidas de tendencia no central, cálculo de frecuencias. Dentro las medidas de dispersión se pueden señalar: la varianza, la desviación típica, el rango.

En los métodos bivariables la herramienta más utilizada son las tablas de contingencia o tabulaciones cruzadas. El contraste que se utiliza para ver si dos variables están relacionadas estadísticamente es la Chi- Cuadrado (χ^2).

Dentro de la inferencia estadística se va a ver: la estimación de parámetros tanto puntual como por intervalos y el contraste de hipótesis, donde nos vamos a centrar en los contrastes paramétricos sobre medias.

4.3. MÉTODOS DE ANÁLISIS UNIVARIABLES Y BIVARIABLES.

➤ MÉTODOS UNIVARIABLES

Entre los métodos univariados se pueden citar los siguientes, los cuales se encuentran divididos entre medidas de posición y medidas de dispersión.

Medidas de posición

1) Medidas representativas de la tendencia central de una distribución:

- Media aritmética.
- Media armónica.
- Media geométrica.
- Media ponderada.
- Mediana y moda.

2) Medidas de posición no centrales:

- Cuartiles.
- Deciles.
- Centiles.
- Percentiles.

3) Otras medidas de posición:

- Frecuencia absoluta y relativa.
- Frecuencia absoluta y relativa acumuladas.

Medidas de dispersión

1) Medidas absolutas:

- Recorrido o rango.
- Rango intercuartílico.
- Varianza.
- Desviación típica.
- Desviaciones absolutas.
- Desviaciones cuadráticas.

2) Medidas relativas:

- Recorrido relativo.
- Coeficiente de variación:
- Basados en desviación absolutas.
- Basados en desviación cuadráticas (coeficiente de Pearson).

➤ **MÉTODOS BIVARIABLES**

Entre los métodos bivariados las tablas de contingencia permiten analizar la información de dos variables simultáneamente. Pero en las tablas de contingencia esas variables tienen que estar medidas con escalas nominales y tener un número limitado de categorías. Las tablas de contingencia van a medir la asociación que existe entre dos variables, pero en ningún caso sirven para detectar relaciones causa-efecto entre las variables.

Ejemplo: Analizar si existe algún tipo de relación entre ir o no de vacaciones y el nivel de ingresos. A continuación, se da la tabla de frecuencias observadas:

Ingresos (A) Irse de vacaciones (B)	REDUCIDOS (A₁)	MEDIOS (A₂)	ELEVADOS (A₃)	<u>Total</u>
SI (B₁)	34	45	55	134
NO (B₂)	53	52	27	132
<u>Total</u>	87	97	82	266

Con observar esta tabla se puede ver que de los 87 con ingresos reducidos más de la mitad no van de vacaciones y que de los 82 con ingresos elevados más de la mitad si van de vacaciones. Por lo tanto, a priori se puede decir que si existe o parecer existir una relación entre ir o no de vacaciones y el nivel de ingresos.

Para ver si esa relación es estadísticamente significativa se va a calcular la Chi-Cuadrado (χ^2) la cual permite comparar dos distribuciones de frecuencia, por una parte, la distribución de frecuencias observadas y, por otra parte, la distribución de frecuencias esperadas que se han de calcular en el caso de que las variables fuesen independientes.

Si A y B son independientes (hipótesis nula) $\Rightarrow P(A_1 \cap B_1) = P(A_1) \cdot P(B_1)$

$$P(A_1) \cdot P(B_1) = \frac{87}{266} \cdot \frac{134}{266}$$

La frecuencia esperada de que suceda ($A_1 \cap B_1$) es igual a: $n \cdot P(A_1 \cap B_1)$.

$$n \cdot P(A_1) \cdot P(B_1) = \left[\frac{87}{266} \cdot \frac{134}{266} \right] \cdot 266 = \frac{87 \cdot 134}{266}$$

En general, la frecuencia esperada se calculará mediante la siguiente expresión:

$$E_{ij} = \frac{n_{Ai} \cdot n_{Bj}}{n}$$

	REDUCIDOS	MEDIOS	ELEVADOS
SI	$87 \cdot 134 / 266 = 43,83$	$97 \cdot 134 / 266 = 48,86$	$82 \cdot 134 / 266 = 41,31$
NO	$87 \cdot 132 / 266 = 43,17$	$97 \cdot 132 / 266 = 48,14$	$82 \cdot 132 / 266 = 40,69$

$$\sum \chi^2 = \frac{(\text{Frecuencias observadas} - \text{Frecuencias esperadas})^2}{\text{Frecuencias esperadas}}$$

$$\chi^2 = \frac{(34 - 43,83)^2}{14,201} + \dots + \frac{(27 - 40,69)^2}{43,83} = 40,69$$

A continuación, se compara el valor de la Chi-Cuadrado con el valor crítico o valor en tablas de una Chi-Cuadrado (χ^2) con:

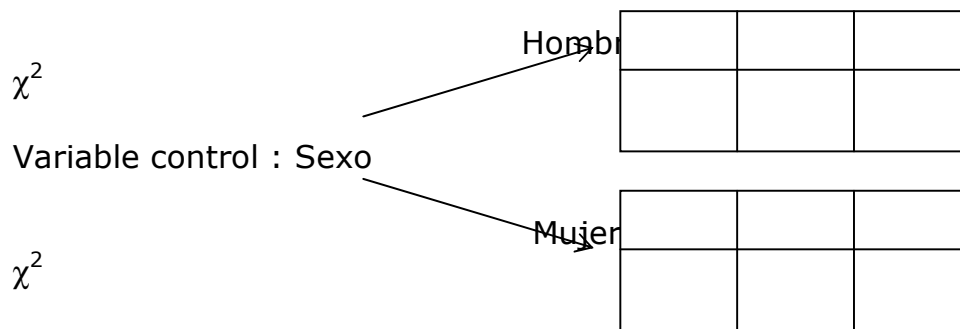
- Grados de libertad : $(n_C - 1) \cdot (n_F - 1) = (3 - 1) \cdot (2 - 1) = 2$ siendo n_C el número de categorías de la variable columna y n_F el número de categorías de la variable fila.
- Nivel de confianza : 95%

En este caso el valor crítico o valor en tablas será igual a: $\chi^2_{(2, 0.05)} = 5,991$. Por lo tanto, se rechaza la hipótesis nula que era que no existía asociación entre las variables, es decir, que las variables A (ingresos) y B (irse de vacaciones) eran independientes. Hay, por lo tanto, una asociación significativa entre las variables.

Para la interpretación de la asociación existente entre dos variables hay que acudir a la tabla de frecuencias observadas, pero también hay que fijarse en las frecuencias relativas, es decir, habrá que calcular porcentajes.

En este caso, interesa los porcentajes por columnas porque son los ingresos los que pueden determinar el que una persona se vaya de vacaciones o no, lo que significa que el ingreso es el que condiciona. Un truco es poner por columnas la variable independiente y luego buscar siempre el porcentaje por columnas.

Pero, ¿cómo se profundiza en el estudio de la asociación existente?. Habrá que ver si la asociación se mantiene cuando se considera variables adicionales. Por ejemplo, en el caso anterior se podría utilizar como variable de control el sexo.



A la asociación original se la conoce como Asociación Total o de Orden Cero. La asociación incluyendo variables de control se llama asociación condicional. Si la asociación desaparece al introducir las variables de control entonces se dice que la asociación inicial era una asociación espúrea (hay una tercera variable, que es la variable de control, que explica a las dos de partida, eso si, explica el comportamiento de las variables iniciales pero no de la asociación inicial).

4.4. MÉTODOS DE INFERENCIA ESTADÍSTICA.

➤ ESTIMACIÓN DE PARÁMETROS

Dentro de la estimación de parámetros se pueden distinguir dos tipos, que serían la estimación de parámetros puntual y la estimación de parámetros por intervalos.

1- Puntual: En la estimación puntual se asigna un valor concreto al parámetro poblacional que se desea estimar. Por ejemplo, utilizar la media muestral como estimador de la media poblacional. Los requisitos que se suelen pedir a un estimador son los siguientes:

- a) Que sea insesgado, lo que quiere decir que las estimaciones que se hagan sean correctas (que estén libres de error).
- b) Que sea consistente, es decir, que a medida que aumenta el tamaño de la muestra (n) mayor sea la probabilidad de que el estimador coincida con el parámetro poblacional (que la estimación sea correcta).

- c) Que tenga la menor varianza posible.
- d) Que sea suficiente, es decir, que utilice el mayor número posible de los datos muestrales disponibles para estimar el parámetro.

2- Por intervalos: Consiste en determinar un intervalo de confianza o un rango de valores entre los que se crea que puede estar el parámetro poblacional a estimar con una probabilidad determinada y alta. Para ello, lo que se hace es, simplemente, sumar y restar al estimador una cantidad concreta que se denomina error máximo. Ese error máximo es igual a:

$$\text{Error máximo} = \frac{S_1}{\sqrt{n}} \cdot Z$$

Z: número de unidades de desviación típica asociadas a un determinado nivel de confianza.

S_1 / \sqrt{n} : es la desviación típica de la distribución muestral del estimador.

El estimador más el error máximo da el límite superior mientras que el estimador menos el error máximo da el límite inferior.

Ejemplo:

Una muestra aleatoria de 1.000 personas responde a una cuestión sobre el gasto mensual en periódicos y revistas observándose una media de 3.200 pesetas y una cuasidesviación típica de 320 pesetas. ¿Entre que límites se encuentra el verdadero gasto en periódicos de la población con un nivel deseado de confianza del 95%?. La pregunta que se nos realiza es sobre cual es la media poblacional, es decir, μ .

n = 1.000 personas
X = 3.200 pesetas
S₁ = 320 pesetas

$$\Rightarrow 3.200 \pm 19,833 \quad = 3.200 \pm 1,96 \frac{320}{\sqrt{1.000}} = 19,833$$

El intervalo que se nos pide será el siguiente: (3180.167, 3219.833). En este intervalo se considera que puede estar la media poblacional con un nivel de confianza del 95%.

➤ **CONTRASTE DE HIPÓTESIS**

El contraste de hipótesis, también denominado prueba de significación, es un método de toma de decisiones que parte de un enunciado o proposición acerca del valor o conjunto de valores que toma un parámetro de la población, y trata de ver o averiguar si se sostiene o se rechaza según los datos muestrales disponibles. Al enunciado o proposición de la que se parte se le denomina "hipótesis nula". Como norma general en marketing a la hora de redactar el enunciado del que se parte se propone el valor del parámetro que entraña un menor riesgo para la empresa que va a tomar la decisión.

La "hipótesis alternativa" son los enunciados que recogen lo que interesa aceptar si se rechaza la hipótesis nula. El error tipo I sería rechazar una hipótesis nula siendo ésta verdadera y el error tipo II sería aceptar una hipótesis nula siendo esta falsa. Estos errores se suelen representar, respectivamente, con α y β .

En la investigación de mercados el procedimiento más común es fijar el nivel de error tipo I que el investigador está dispuesto a aceptar, es decir, α es la probabilidad de cometer el error tipo I. El nivel de error con el que se suele trabajar es de 0.05, por lo tanto, $1 - \alpha$ es igual a 0,95. Al α también se le denomina nivel de significación y a $1 - \alpha$ se le llama nivel de confianza.

Nos vamos a centrar en los contrastes paramétricos sobre medias, en concreto en tres casos, que son:

- a) Contraste de hipótesis para la media de una muestra.
- b) Contraste de hipótesis para las medias de dos muestras independientes.

- c) Contraste de hipótesis para las medias de muestras relacionadas.

Se va a asumir que hay normalidad y que se está trabajando con escalas de intervalo o de razón.

a) Contraste de hipótesis para la media de una muestra: Permite comparar la media muestral con una hipotética media poblacional, y ver si hay diferencias significativas entre ellas. Hay dos tipos de pruebas que se pueden realizar para llevar a cabo la comparación.

- Prueba z: esta prueba se utiliza cuando la desviación típica poblacional es conocida o cuando no se conoce la desviación típica poblacional y el tamaño de la muestra es mayor de 30.
- Prueba t: esta prueba se utiliza cuando no se conoce la desviación típica poblacional y el tamaño de la muestra es menor de 30.

La prueba que se utiliza con una mayor frecuencia es la prueba z ya que no se suele conocer la desviación típica poblacional y el tamaño de la muestra suele ser mayor de 30. La expresión a utilizar sería la siguiente:

$$z = \frac{\bar{X} - \mu}{S_1 \sqrt{n}}$$

Luego se compara el valor de z con el valor crítico o valor en tablas para un determinado nivel de confianza. Si z es mayor que el valor crítico para un nivel de confianza se rechaza la hipótesis nula.

Ejemplo:

Una empresa industrial está interesada en conocer si una nueva máquina de embotellado recibe una valoración superior a 7 en una escala de intención de compra de diez puntos. Para ello contacta con una muestra de 40 ingenieros responsables de compra de otras tantas plantas embotelladoras a quienes se les explica el funcionamiento del nuevo producto. Los resultados obtenidos son: media muestral igual a 7.9, cuasidesviación típica de 1,6. ¿Será

introducido ese nuevo producto en el mercado para un nivel de confianza del 95%?

La hipótesis nula para este problema es que $\mu \leq 7$ y la hipótesis alternativa sería que $\mu > 7$. Para un nivel de confianza del 95% el valor crítico será igual a 1,64. En este caso se utiliza la prueba z porque el tamaño de la muestra (n) es igual a 40 que es mayor de 30 y porque la desviación típica poblacional es desconocida.

$$z = \frac{\bar{X} - \mu}{S_1 \sqrt{\frac{1}{n}}} = \frac{7,9 - 7}{\sqrt{1,6 / 40}} = 3,5576$$

Se rechaza la hipótesis nula porque z es mayor que el valor crítico (3,55 > 1,64) y, por lo tanto, se lanza el nuevo producto al mercado porque la intención de compra va a estar por encima de 7 en una escala de 10 puntos.

b) Contraste de hipótesis para las medias de dos muestras independientes: Nos permite comparar las medias muestrales de dos grupos de individuos distintos que han sido seleccionados aleatoriamente. Como en el caso anterior hay dos posibles pruebas que son:

- Prueba z: esta prueba se utilizará si la desviación típica poblacional es conocida o si ésta no es conocida y el tamaño de la muestra es mayor de 30.

- Prueba t: se utiliza esta prueba si la desviación típica poblacional no es conocida y el tamaño de la muestra es menor de 30.

Además de fijarnos si el tamaño de la muestra es mayor o menor de 30 y si la desviación típica poblacional es conocida o desconocida habrá que tener en cuenta la homogeneidad de varianzas. En el caso de que haya homogeneidad de varianzas se utilizará una expresión determinada y si no hay homogeneidad se utilizará otra expresión que sería el test de Levene. La expresión a utilizar en el caso de que la desviación típica poblacional sea desconocida, el tamaño de la muestra sea mayor de 30 y haya homogeneidad de varianzas será la siguiente:

$$z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

La hipótesis nula (H_0) suele ser la siguiente: $\mu_1 = \mu_2$. Mientras que la hipótesis alternativa (H_1) suele ser la siguiente: $\mu_1 \neq \mu_2$.

El resultado obtenido mediante la anterior expresión se compara con el valor crítico para un determinado nivel de confianza y si z es mayor que el valor crítico se rechaza la hipótesis nula.

Ejemplo:

Un fabricante de patatas fritas decide realizar un test de prueba para un nuevo producto que son las patatas fritas al ajo. Una muestra de 100 personas menores de 30 años le otorga una valoración media de 4,9 con una cuasidesviación típica de 1,7 en un test de sabor que compara dicho nuevo producto con la elaboración de patatas fritas tradicionales. El valor medio de otra muestra de 100 personas mayores de 30 años es de 4,3 con una cuasidesviación típica de 1,8. ¿Se puede considerar diferente la opinión de ambos segmentos de mercado para un coeficiente o nivel de confianza del 95%?.

En este caso se utiliza la prueba Z porque el tamaño de la muestra es mayor de 30 y no se conoce la desviación típica poblacional y se va a asumir la homogeneidad de varianzas. La hipótesis nula será que no hay diferencias entre las opiniones de los menores de 30 años y las de los mayores de 30 años ($\mu_1 = \mu_2$). La hipótesis alternativa será que si hay diferencias entre las opiniones de los dos segmentos de mercado ($\mu_1 \neq \mu_2$).

$$z = \frac{4,9 - 4,3}{\sqrt{\frac{(100 - 1) \cdot 1,7^2 + (100 - 1) \cdot 1,8^2}{100 + 100 - 2} \cdot \frac{100 + 100}{100 \cdot 100}}} = 2,423$$

Para un nivel de confianza del 95% el valor crítico es igual a 1,96. El valor que se obtiene mediante la anterior expresión es mayor que el valor crítico de forma que se rechaza la hipótesis nula, lo que quiere decir que hay diferencias entre las opiniones, lo que llevaría a segmentar el mercado y adaptar la política de comunicación para dirigirla hacia los jóvenes.

d) Contraste de hipótesis para las medias de muestras relacionadas: Tenemos un mismo grupo de entrevistados que valoran un producto o servicio en dos momentos distintos del tiempo, que valoran dos productos competidores, o que valoran la importancia de dos atributos o aspectos distintos de un servicio o producto.

En estas situaciones se crea una variable nula igual a la diferencia de opiniones. La hipótesis nula va a ser que no existe diferencia entre las opiniones, valoraciones o puntuaciones. Las pruebas que se pueden utilizar son las mismas que se utilizan en los dos casos anteriores. La expresión a utilizar, en el caso de que el tamaño de la muestra sea mayor de 30 y la desviación típica poblacional sea desconocida, será la siguiente:

$$z = \frac{D^-}{S_D \sqrt{n}}$$

A continuación, se compara el valor obtenido con esta expresión con el valor crítico para un determinado nivel de confianza, y si z es mayor que el valor crítico entonces se rechaza la hipótesis nula.

Ejemplo:

Un investigador de mercados selecciona una muestra aleatoria de 270 personas quienes valoran la importancia de dos atributos a la hora de seleccionar entidades bancarias, siendo dicho atributos los siguientes:

- La amabilidad o atención personal.
- La rentabilidad ofrecida por las inversiones.

Los resultados obtenidos son los siguientes:

	n	\bar{D}	S_D
Amabilidad	270	5,63	0,67
Rentabilidad	270	4,19	1,62
Diferencias en las opiniones para cada individuo	270	1,44	1,61

¿Hay evidencia suficiente para afirmar, con un nivel deseado de confianza del 95%, que la importancia asignada a ambos atributos es diferente?. Se utiliza la prueba z porque el tamaño de la muestra es mayor de 30 y la desviación típica poblacional es desconocida. La hipótesis nula (H_0) sería que no existen diferencias en las opiniones, es decir, la diferencia entre la importancia de la amabilidad y la rentabilidad es nula.

$$z = \frac{1,44}{1,61 \sqrt{270}} = 14,6967$$

En este caso se rechaza la hipótesis nula porque Z es mayor que el valor crítico para un nivel de confianza del 95% ($14,69 > 1,96$). Desde el punto de vista del marketing esto supone que si hay diferencia en la importancia que se concede a los distintos atributos, amabilidad y rentabilidad, asignando una mayor importancia a la amabilidad.

Referencias bibliográficas.

- Artículo de revista: Pérez Díaz, J.L. (2000): "Título del Artículo", Revista, Volumen 54, Número 2, Páginas 41 – 53.
- Libro: Pérez Díaz, J.L. (2000): "Título del Libro", Editorial Díaz de Santos, Madrid (7ª edición).
- STANTON, William. Fundamentos de Marketing. McGraw-Hill. Décima Primera Edición. Méjico 2000 (únicamente el capítulo 4)
- POPE, Jeffry. Investigación de Mercados. Editorial Norma. 1986
- LOPEZ ALTAMIRANO, Alfredo. ¿ Qué son, para qué sirven y cómo se hacen las Investigaciones de Mercado? Editorial CECSA. Primera Edición. Méjico 2001.
- Kotler, P "Dirección de la Mercadotecnia, Análisis, Planeación, Implementación y Control" sexta edición 1992.
 - - Lambin, JJ "Marketing Estrategico" Ed. Mc Graw-Hill
 - Ceavens, Hills, Woodruff Administración en mercadotecnia Ed. CECSA
 - Andrés Quijano Ponce de León, Elementos básicos de la mercadotecnia
 - Gist, Ronald R., Principio de Mercadotecnia, Editora Interamericana, México, 1973,448 págs.
 - Thomas, Kinnear, Investigación de Mercado, MCGraw-Hill, Colombia, 1994, 760 págs.