

## CAPÍTULO III

### Metodología

El presente capítulo tiene la finalidad de explicar el procedimiento que se llevó a cabo para realizar la investigación, es decir, se dará una descripción generalizada de los sujetos que se estudiaron así como sus características, además del material empleado para realizar la investigación, y de los lugares donde se llevo a cabo la recolección de datos.

#### 3.1. Tipo De Investigación

De los cuatro tipos de investigación que existen: exploratorios, descriptivos, causales y predictivos, es el descriptivo el más apropiado en este caso; ya que una investigación de tipo descriptivo o correlacional, como también se le conoce, es aquella que, se propone describir las características de ciertos grupos, determinar la frecuencia de algunos sucesos, evaluar la relación entre dos o más variables o realizar predicciones (Santesmases, 2001).

A su vez, la investigación descriptiva se divide en estudio longitudinal y estudio transversal. El estudio longitudinal consiste en realizar mediciones repetidas de un mismo hecho con la finalidad de mostrar el cambio en el comportamiento de las variables estudiadas. El estudio transversal, o de sección cruzada, tiene la finalidad de proporcionar una “fotografía” de las variables que se pretenden estudiar así como la relación entre ellas en un momento determinado (Santesmases, 2001). Por lo tanto es el estudio transversal el que más se adapta a las necesidades de esta investigación.

El estudio transversal se puede llevar a cabo mediante un estudio en profundidad o encuestas. Las encuestas son un tipo de investigación cuantitativa muy apropiado para recabar información en una investigación social o de mercados. Generalmente se basa en grandes muestras representativas de la población que se estudia. Los datos se obtienen mediante comunicación a través de entrevistas personales y utilizando un cuestionario (Santesmases, 2001).

### **3.2. Selección De La Muestra**

“Los elementos de una población son las unidades de las que se busca la información; son los individuos, las unidades elementales que forman la población acerca de la cual se va a hacer inferencias” (Kish, 1975, p. 27).

Para el estudio se realizó un muestreo por cuotas, el cual es un muestreo no probabilístico, esto es, que el proceso de selección de los elementos de la muestra es no aleatorio y se realiza total o parcial según criterios del investigador. El muestreo por cuotas consiste en seleccionar la muestra de manera que sus características se adecuen a los controles o *cuotas* establecidos por el investigador. Estas características pueden ser género, edad, nivel de ingresos, lugar de residencia, entre otros (Santesmases, 2001).

Santesmases (2001) menciona:

La determinación del tamaño de la muestra está en función del presupuesto y tiempo disponible, así como del error que se admita en los datos obtenidos. Este error que se comete al tomar una muestra en lugar de la totalidad de la población se denomina

**error de muestreo** y es fácilmente cuantificable cuando el muestreo es probabilístico.

Para la elección de los sujetos a estudiar se partió de la información que brinda el INEGI sobre la población de la Ciudad de Puebla en el XII Censo General de Población y Vivienda 2000. Con base en lo anterior se determinaron las siguientes cuotas: edad y género. Los grupos de edades se pueden observar en el apéndice 1. Cabe mencionar que para efecto de esta investigación se reunieron los datos en grupos quinquenales, a partir de 18 años en adelante (apéndice 2).

Para determinar el tamaño de la muestra se llevó a cabo una prueba piloto en la Universidad de las Américas, Puebla, la cual constó de 15 encuestas, y el resultado obtenido fue que el tiempo promedio de aplicación por encuesta es de 6 minutos. Calculando el tiempo disponible para la aplicación de las encuestas, se determinó que la muestra constaría de 230 personas.

Las entrevistas fueron levantadas en plazas comerciales (Plaza Dorada y La Noria), en universidades a personas residentes en la Ciudad de Puebla (Universidad de las Américas, Puebla, Universidad Popular Autónoma del Estado de Puebla, Universidad Tecnológica de Puebla) y en el Centro Histórico de la Ciudad de Puebla.

### **3.3. Material**

El material que se utilizó para esta investigación es un cuestionario (apéndice 4). Este se define como un formulario que contiene preguntas diseñadas con el propósito de recabar

los datos necesarios para lograr los objetivos que busca la investigación (McDaniel, Gates, 1999). El cuestionario empleado consta de preguntas disyuntivas, que son aquellas en las que el encuestado tiene como posible respuesta un sí o no o un falso o verdadero; de opción múltiple, siendo éstas aquellas preguntas en las que se presentan varias alternativas específicas como respuesta y; abiertas, que son en las que el encuestado tiene plena libertad para expresar lo que piensa sin limitación alguna. Siendo la finalidad de estas servir como herramienta para indagar el grado de conocimiento que tiene la muestra acerca del té y todo lo relacionado con éste, además de las preferencias y el nivel de consumo de esta bebida.

#### **3.4. Procedimiento**

Para la aplicación de las encuestas se llevó a cabo un muestreo por cuotas. Para las encuestas se buscaron personas que estuvieran dentro de los controles definidos anteriormente.

Al encuestado se le explicó la finalidad de la encuesta. Después el encuestador hizo las preguntas de manera oral y él mismo capturó los datos en los formatos impresos de las encuestas. Al finalizar la encuesta, el encuestador agradecía la colaboración al encuestado. Durante la aplicación de la encuestas se sumó el número de personas de cada cuota para verificar que se estuvieran llevando a cabo las ponderaciones (apéndice 3).

Una vez terminada la aplicación de las encuestas se procedió al análisis estadístico de los datos obtenidos con la ayuda de software especializado.

### **3.5. Análisis De Datos**

El software aplicado para el análisis de datos fue StatView, el cual es empleado para llevar a cabo análisis estadísticos. Este software además de realizar un análisis simple de la información obtenida, permitió cruzar dos y más datos con el objeto de llegar a un análisis más consistente.

Para llevar a cabo el análisis estadístico se agruparon las posibles respuestas en categorías para facilitar la captura y el análisis de los datos.

Los análisis que se realizaron fueron histogramas de frecuencias de cada una de las preguntas del cuestionario, tablas de contingencia entre dos o más variables.

#### **3.5.1 Histogramas De Frecuencia**

Los histogramas de frecuencia es un modo sencillo y útil de presentar los datos de tal manera que las diferencias entre frecuencias se destaquen fácilmente, para ello se sirve de figuras de áreas o alturas proporcionales a las frecuencias en cada categoría (Blalock, 1978). Los histogramas de frecuencia son representaciones gráficas de los datos obtenidos en un diagrama de columnas para una mejor comprensión de la información (Hawkins, 1980).

#### **3.5.2 Tablas De Contingencia**

Las tablas de contingencia son tablas que constan de columnas y renglones, los cuales “totalizan” la clasificación cruzada de elementos de la muestra de acuerdo a dos características. En estas, el investigador esta interesado en saber si hay una relación entre

las dos características; una representada por los renglones y, la otra, por las columnas de la tabla. La hipótesis nula consiste en que las dos características son independientes (Hawkins, 1980).

Las tablas de contingencia son empleadas para conocer la independencia y homogeneidad entre dos variables. Para ello se plantea la hipótesis nula que se ha de aceptar si el valor que se obtuvo en el análisis en el Chi Square P-Value es menor al valor de alfa 0.1 y rechazar si el valor de alfa es mayor a 0.1. Un valor de alfa se refiere a la probabilidad de rechazar la hipótesis nula cuando ésta es en realidad verdadera, esto se conoce normalmente como error tipo I (Blalock, 1978). Los posibles valores de significación son normalmente 0.05, 0.01 y 0.1, se escogió el valor de 0.1 puesto que es un valor que da un poco más de holgura para realizar los análisis y obtener conclusiones.

Las tablas de contingencia cuentan con los siguientes datos:

- Num. Missing, que es el total de datos faltantes en la base de datos en relación con las variables empleadas en dicha tabla;
- DF, son los grados de libertad, se define bajo la fórmula:  $df=(r-1)(c-1)$ , donde  $r$  son los renglones y  $c$  las columnas;
- Chi Square, que es una prueba que se emplea para apreciar si las frecuencias obtenidas son diferentes o no de las frecuencias que se esperaban bajo supuestos teóricos, su fórmula es:  $X^2 = \sum (f_o - f_e)^2 / f_e$ ; en lo que  $f_o$  y  $f_e$  se refieren respectivamente a las frecuencias observadas y esperadas para cada casilla de la tabla de contingencia;
- Chi Square P-Value, que es la probabilidad de observar lo que se observó si la hipótesis nula es cierta. Para efectos de este estudio, el dato relevante es el

Chi Square P-Value, para poder verificar si la hipótesis nula es aceptada o no.

- G-Squared, G-Squared P-Value, Cramer's V, estos, son pruebas de hipótesis, que aparecen en las tablas. Para efectos del presente estudio no serán utilizados.
- Coeficiente de Contingencia, es aquel valor que proporciona la relación que hay entre las variables que se estudian, entre más cerca del uno, más estrecha es la relación entre las variables.
- La prueba de Fisher, ésta proporciona probabilidades exactas y no aproximadas, en las tablas de contingencia que son de dos por dos celdas, este valor es el que se debe comparar con el valor de alfa escrito anteriormente (Blalock, 1978).

Es importante mencionar que la edad de los encuestados se agrupó en “Joven”, todos aquellos que tienen de 18 a 28 años; “Adulto Joven”, cuando su edad es de entre los 29 y 45 años; “Adulto Mayor”, todos aquellos mayores de 45 años.