CHAPTER 5: METHODOLOGY

5.1. SAMPLING

In order to achieve the above-mentioned objectives data was collected from primary sources, that is, the consumers. In this research we have used convenience as well as judgement sampling.

Convenience sampling is used in exploratory research where the researcher is interested in getting an inexpensive approximation of the truth. As the name implies, the sample is selected because they are convenient. This non probability method is often used during preliminary research efforts to get a gross estimate of the results, without incurring the cost or time required to select a random sample. Judgment sampling is a common non probability method. The researcher selects the sample based on judgment. This is usually an extension of convenience sampling. For example, a researcher may decide to draw the entire sample from one "representative" city, even though the population includes all cities. When using this method, the researcher must be confident that the chosen sample is truly representative of the entire population. Shoppers who were shopping in lifestyle retail outlets were targeted to gather the necessary information. The brands in consideration were:

Private Label Brands
Hush,Fig, rio avasa netplay net work, dnmx, team spirit, performax, Unisex sports, Frendz, Atra, Erato, Frescolino, Engine, Abhinandan, Elaap, Dxi, Forever fashion, Agno3, Zaira, Kajali, Purple nasty, Melange, Strings, Fame forever, Code, Ucla, Ginger, Smiley, Forca, Juniors, Kappa, Bossini, Bossini kids, Milange, Kashish, Haute curry, Sanaa, Austin reed, Elizza donatein, Life Vetorio fretini, Acropolis, Stop, Honey, Ajile, Umm, Rig, Bare, Rangmanch Trosha, Akriti, Annabelle, Chirpie pie, Chalk kids, Bare kids, Rig kids, John miller, Bare, dj&c, Indigo nation, Rig.
National brands Allen solly, Van huesen, Louis phillip, Numero uno, Levis, Lee, lee Cooper, Jockey, SDL, Prefull, Bonjour, Adidas, Lilliput, Gini jony, Chabra

5.2. RESEARCH INSTRUMENT

The questionnaire was developed to directly address the goals of the study. Long questionnaires get less response than short questionnaires. So our questionnaire was kept short.

They were very easy to understand, short sentences and basic vocabulary was used. We used simple and direct language as the questions needed to be clearly understood by the respondent. The wording of questions was simple and to the point we did not use uncommon words or long sentences. Items were made as brief as possible. This helped in reducing misunderstandings and made the questionnaire appear easier to complete.

Adequate space was left for respondents to make comments as we wanted the respondent to complete our questionnaire. An attempt was made to make questionnaire interesting by providing variety in the type of items used. Varying the questioning format also prevented respondents from falling into "response sets". At the same time, it was important to group items into coherent categories so that all items should flow smoothly from one to the next.

Each question followed comfortably from the previous question. Transitions between questions was smooth as the Questions that jump from one unrelated topic to another feel disjointed and are not likely to produce high response rates.
5.3. SAMPLE COLLECTION

The questionnaire was used along with personal interview for collecting data from the shoppers. Retail outlets like Shoppers Stop, Reliance trends, Lifestyle, Westside Big Bazaar, CnM, Max were targeted as these stores have a strong presence in Private Label brand.

The main factors to be considered in this task were:

- The seriousness of the respondent
- The completion of the questionnaire by the respondent
- Choice of right mix of respondents based on
  1. Age
  2. Income
  3. Gender
  4. Location

The final questionnaire was offered to 599 respondents at various locations mentioned below.

1. Lifestyle (Rohini)
2. CnM (Pitam Pura)
3. Big Bazaar (Netaji Subaash Place)
4. Ritu Wears (Rohini)
5. Pantaloons (Rithala)
6. Max (Rajouri Garden)
7. Shopper Stop (Rohini)
8. Reliance Retail (Rohini)
9. Globus (Noida)
10. Westside (Rajouri Garden)
The above mentioned outlets are the most popular outlets of Delhi where there is a mix of shoppers as we desire in our research. The outlets have good variety of private label brands as well as leading national brands.

The questionnaires were filled through in depth personal interview with the respondent in face to face situation. This served to provide maximum control over the interviewing process and reduced the variability of result caused by differences in respondent and interviewer characteristics. The desired questions were asked and answer obtained. The recording of information was done sometimes during the interview sometimes after the interview.

The process of getting the questionnaire filled and understanding the responses by the people I interacted with has provided me with in-depth understanding of the attitude and perception of the Indian shopper. The background of literature review provides solid foundation for the analysis and quick understanding of the perceptions of the shopper.

5.4. ANALYSIS OF DATA

Analysis of data is a process of inspecting, cleaning, transforming, and modelling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

The filling in of questionnaire by personal interview method resulted in generation of large quantum of qualitative data. The qualitative data so collected was subject to logical and structured process of analysis.
Substantial time and effort were invested in absorbing and understanding the interview transcripts. This involved repeated reading of transcripts running into thousand sheets over a period of nearly thirteen weeks, which was conducted in varied settings in order to maximize absorption of responses generated.

Additionally any significant comments relating to the topic were also captured. Where relevant, the data was validated for logical and consistency.

The findings distilled from above data were categorised based on common themes, which harmonized with the areas being investigated in the study

5.4.1. DATA PREPARATION

Data Preparation involved checking or logging the data in; checking the data for accuracy; entering the data into the computer; transforming the data; and developing and documenting a database structure that integrates the various measures

5.4.2. LOGGING THE DATA

Initially my task was to set up a procedure for logging the information and keeping track of it until I am ready to do a comprehensive data analysis. I wanted to set up a database that enables anyone to assess at any time what data is already in and what is still outstanding. I fulfilled this objective with standard statistical programs SPSS.

A database for logging incoming data is a critical component in good research record-keeping. After getting my questionnaires filled from 700 respondents I tabulated the data on Microsoft excel sheets.
5.4.3. CHECKING THE DATA FOR ACCURACY

As soon as data was received I screened it for accuracy. There are several questions which I asked myself as part of this initial data screening:

- Are the responses legible / readable?
- Are all important questions answered?
- Are the responses complete?
- Is all relevant contextual information included (e.g., data, time, place)?

In most market research, quality of measurement is a major issue. The process of Assuring that the data collection process does not contribute inaccuracies helped me to assure the overall quality of subsequent analyses.

Interestingly only 599 cases were proved thoroughly accurate. Thus resulting in reducing my sample size to 599 respondents.

5.4.4. DEVELOPING A DATABASE STRUCTURE

The database structure is the manner in which you intend to store the data for the study so that it can be accessed in subsequent data analyses. You might use the same structure you used for logging in the data or, in large complex studies; you might have one structure for logging data and another for storing it. There are generally two options for storing data on computer -- database programs and statistical programs. Usually database programs are the more complex of the two to learn and operate, but they allow the analyst greater flexibility in manipulating the data.

In my research project, I generated a printed codebook that described the data and indicated where and how it can be accessed.
The codebook included the following items for each variable:

- Variable name
- Variable description
- Variable format (number, data, text)
- Instrument/method of collection
- Date collected
- Respondent or group
- Variable location (in database)
- Notes

The codebook is an indispensable tool for the analysis. Together with the database, it provided me comprehensive documentation that enables other researchers who might subsequently want to analyze the data to do so without any additional information.

5.4.5. ENTERING THE DATA INTO THE COMPUTER

There are a wide variety of ways to enter the data into the computer for analysis. Probably the easiest is to just type the data in directly. That is exactly how I entered my data this provided control and accuracy over whole process.

5.4.6. DATA TRANSFORMATIONS

Once I had entered the data it became necessary to transform the raw data into variables that are usable in the analyses.

There are a wide variety of transformations that I had to perform. Some of the more common were:

- Missing values
Many analysis programs automatically treat blank values as missing. I had to check the specific program (SPSS) I was using to determine how to handle missing values.

- Categories

For many variables I had to collapse them into categories. For instance, I had to collapse income estimates into income ranges, Occupation into occupation ranges.

Finally I started with applying statistical techniques to my data.

5.5. STATISTICAL TOOLS

5.5.1. Mean

This may be the most average definition of average (whatever that means). This is the weighted average—a total of all numbers included divided by the quantity of numbers represented. The most common expression for the mean of a statistical distribution with a discrete random variable is the mathematical average of all the terms. To calculate it, add up the values of all the terms and then divide by the number of terms. This expression is also called the arithmetic mean. There are other expressions for the mean of a finite set of terms but these forms are rarely used in statistics.

The mean of a statistical distribution with a continuous random variable, also called the expected value, is obtained by integrating the product of the variable with its probability as defined by the distribution.

5.5.2. Standard deviation
To overcome the problem of dealing with squared units, statisticians take the square root of the variance to get the standard deviation.

The standard deviation (for a sample) is defined symbolically as

\[ s = \sqrt{\text{var}} = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} \]

So if the scores in the data were 5, 7, 6, 1, and 8, their squared differences from the mean would be 0.16 (from \([5-5.4]^2\)), 2.56 (from \([7-5.4]^2\)), 0.36 (from \([6-5.4]^2\)), 19.36 (from \([1-5.4]^2\)), and 6.76 (from \([8-5.4]^2\)). The mean of these squared deviations is 5.84 and its square root is 2.41 (if dividing by N), which is the standard deviation of these scores. The standard deviation is defined as the average amount by which scores in a distribution differ from the mean, ignoring the sign of the difference. Sometimes, the standard deviation is defined as the average distance between any score in a distribution and the mean of the distribution.

The above formula is the definition for a sample standard deviation. To calculate the standard deviation for a population, N is used in the denominator instead of N-1. Suffice it to say that in most contexts, regardless of the purpose of your data analysis, computer programs will print the result from the sample sd. So we will use the second formula as our definitional formula for the standard deviation, even though conceptually dividing by N makes more sense (i.e., dividing by how many scores there are to get the average). When N is fairly large, the difference between the different formulas is small and trivial. Using the N-1 version of the formula, we still define the standard deviation as the average amount by which scores in a distribution differ from the mean, ignoring the sign of the difference, even though this is a true average using this formula.
68% of the values are within one standard deviation of the mean. 95% of the values are within two standard deviations of the mean. More than 99% of the values are within three standard deviations of the mean.

5.5.3. Paired t test

A paired t-test compares two samples in cases where each value in one sample has a natural partner in the other. A paired t-test looks at the difference between paired values in two samples, takes into account the variation of values within each sample, and produces a single number known as a t-value.

You can find out how likely it is that two samples from the same population (i.e. where there should be no difference) would produce a t-value as big, or bigger, than yours. This value is called a p-value. So, a t-test measures how different two samples are (the t-value) and tells you how likely it is that such a difference would appear in two samples from the same population (the p-value).

Most software packages report the t-value as part of the summary of the results of a t-test. The value of t is the result of putting the sample data through the formula for the t-test.

The t-value is related to the size of the difference between the means of the two samples you are comparing. The larger t is, the larger the difference.

Here is the formula for a paired t-test.

$$
t = \frac{\sum d}{\sqrt{\frac{n\left(\sum d^2\right)-\left(\sum d\right)^2}{n-1}}}
$$

The top of the formula is the sum of the differences (i.e. the sum of $d$). The bottom of the formula reads as:
The square root of the following: \( n \) times the sum of the differences squared minus the sum of the squared differences, all over \( n-1 \).

- The sum of the squared differences: \( \sum d^2 \) means take each difference in turn, square it, and add up all those squared numbers.
- The sum of the differences squared: \((\sum d)^2\) means add up all the differences and square the result.

Brackets around something in a formula mean (do this first), so \((\sum d)^2\) means add up all the differences first, then square the result.

5.5.4. Duncan MRT

Duncan's multiple range tests, or Duncan's test, or Duncan's new multiple range test, provides significance levels for the difference between any pair of means, regardless of whether a significant \( F \) resulted from an initial analysis of variance. It is a common statistical technique for determining if differences exist between three or more groups. Duncan's new multiple range test (MRT) is a variant of the Student–Newman–Keuls method that uses increasing alpha levels to calculate the critical values in each step of the Newman–Keuls procedure. Duncan's MRT attempts to control family wise error rate (FWE) at \( \alpha_{ew} = 1 - (1 - \alpha_{pc})^{k-1} \) when comparing \( k \), where \( k \) is the number of groups. This results in higher FWE than unmodified Newman–Keuls procedure which has FWE of \( \alpha_{ew} = 1 - (1 - \alpha_{pc})^{k/2} \).

David B. Duncan developed this test as a modification of the Student–Newman–Keuls method that would have greater power. Duncan's MRT is especially protective against false negative (Type II) error at the expense of having a greater risk of making false positive (Type I) errors.