

Chapter 3

RESEARCH METHODOLOGY

The present chapter discusses the research design and the detailed research methodology followed to undertake the current study. It consists of a discussion of methods of sample selection, identification of variables to be studied, data extraction, and finally the techniques of data analysis.

3.1 Research Design

The research design refers to the overall strategy that one adopts to integrate the different components of the study in a coherent and logical way, thereby, ensuring that the research problem is addressed effectively; it constitutes the blueprint for the collection, measurement, and analysis of data. According to Churchill (2010), there are three major research designs namely; Exploratory, Descriptive and Causal or Experimental. In Exploratory research the “major emphasis is on the discovery of ideas and insights” (p. 128). Descriptive Research is used to obtain information concerning the current status of the phenomena and to describe "what exists" with respect to variables or conditions in a situation (libguides.usc.edu) and Causal research design is majorly concerned with determining “*cause and effect* relationships” (Churchill, 2010, p. 128).

Descriptive research design is chosen for the present study owing to its unique characteristics that are best suited for the requirements of this research work.

Descriptive research designs help provide answers to the questions of who, what, when, where, and how, associated with a particular research problem. This falls in line with the objective of the present study where the researcher describes the style of the given biopics in

an attempt to understand and identify the similar patterns of style and structure in these biopics.

Descriptive research does not fit neatly into the definition of either quantitative or qualitative research methodologies, but instead it can utilize elements of both, often within the same study. It involves gathering data that describe events and then organizes, tabulates, depicts, and describes the data collection (Glass & Hopkins, 1984). It also includes collection of quantitative information that can be tabulated along a continuum in numerical form. This study extracts both qualitative and quantitative information from the sample units and uses the coded data in tabular form for further analysis.

Unique in the number of variables employed, descriptive research like other types of research, can include multiple variables for analysis, yet unlike other methods, it requires only one variable (Borg & Gall, 1989). This helps in analyzing either single variable or multiple variables to identify the desired pattern.

Moreover, along with the extensive use of descriptive methods and analysis techniques that yield reports concerning the measures of central tendency, variation, and correlation; it can employ methods of analyzing correlations between multiple variables by using tests such as Pearson's correlation, regression, or multiple regression analysis. The combination of its characteristic summary and correlational statistics, along with its focus on specific types of research questions, methods, and outcomes is what distinguishes descriptive research from other research types.

As descriptive research uses both quantitative and qualitative methodologies, it enables the researcher to describe events in greater or less depth as needed, to focus on various elements of different research techniques, and to engage quantitative statistics to organize information in meaningful ways. Three main purposes of research are to describe, explain, and validate findings by creative exploration which serves to organize the findings in order to align them

with explanations, and then test or validate those explanations (Krathwohl, 1993). Like many descriptive studies that describe the natural or man-made phenomena such as their form, structure, activity, and change over time, relation to other phenomena, and so on; this study intends to gain insight into the form and structure of biopics through the description of their style. Although, the present study is more concerned with the description and explanation of findings rather than their validation, the design for the study remains largely descriptive.

Descriptive research is often used as a pre-cursor to more quantitative research designs with the general overview giving some valuable pointers as to what variables are worth testing quantitatively. If the limitations are understood, they can be a useful tool in developing a more focused study. Descriptive studies can yield rich data that leads to important recommendations in practice.

3.2 Procedure for Sample Selection

Once the problem is clearly specified and an appropriate research design is developed, an important step in the research process is to select those elements from which the information will be collected. This section describes in detail the six- step sampling procedure given by Churchill, 2010, which is as follows: 1) Defining the population, 2) Identifying the sample frame, 3) Selecting a sampling procedure, 4) Determining the sample size, 5) Selecting the sample elements and 6) Collecting the data from the designated elements.

3.2.1 Population

Population is defined as the “totality of the cases that conform to some designated specifications.” The specifications define the elements that belong to the target group and those that are to be excluded (Churchill, 2010, p. 535). The target population, thus, is “the collection of elements or objects that possess the information sought by the researcher and about which inferences have to be made” (Malhotra and Dash, 2011, p. 330).

As the present study deals with an analysis of biopics, population consists of films which are selected according to the following criteria:

1. While deciding upon which film can be considered a biopic, the definition given by Custen (1992) is considered, which says that a biopic “is minimally composed of the life, or the portion of the life, of a real person whose real name is used” (p. 6). Defining a biopic has helped in removing those films from the list which are loosely based on a real life but refrain from any claims to authenticity by avoiding the use of real names.
2. Only those biopics are considered which are released in Hindi and are released during the period 2011 to 2016. The two basic reasons for the selection of this time frame are as follows: Firstly, in order to consider the latest five year span for the study and Secondly, the Indian cinema witnessed a recent surge in the production of biopics, especially after 2010 with a number of biopics being commercially and critically successful. Also, a major shift is evident in the approach to biopic filmmaking as filmmakers are taking up common people as biopic subjects which has redefined the growth and development of the genre in Indian Cinema. This period of five years thus offers an interesting time frame for research.
3. The biopic is a complex genre due to its hybrid nature. As it shares characteristics with other genres, often a biopic is simultaneously categorized in two or more genres. To simplify the selection procedure, *imdb* category classifications are used. As *imdb* classifies a particular film into all the possible categories like, drama, thriller, biography, horror, etc. therefore, only those biopics are considered which have one of the classifications as ‘biography’. (Cutting et.al, 2011, p. 570)

4. Only those biopics are considered which have a theatrical release and have followed standard release procedure¹ (Wook Ji and Waterman, 2010, p. 8). The biographical films which are not released according to this procedure are not considered.

In order to obtain a comprehensive list of all the Indian biographical films produced during the selected period, online databases have been explored. This includes National Film Archive of India (NFAI), imdb.in and Wikipedia and then the combined list from all the relevant sources excluding the repetitions have been compiled.

3.2.2 Sample Frame

Identifying the sample frame includes the listing of the elements from which the actual sample is drawn. For the present study, after defining the population, a sample frame of sixteen films is obtained and the final sample of five films is drawn from it.

3.2.3 Sampling Procedure

Sampling techniques can be broadly classified as Non-probability and Probability. Non-probability sampling relies on the personal judgment of the researcher rather than chance to select sample elements. The researcher can arbitrarily or consciously decide what elements to include in the sample. Probability sampling on the other hand is the procedure in which each element of the population has a fixed probabilistic chance of being selected for the sample (Malhotra and Dash, 2011). Non-probability sampling consists of convenience, judgment and quota sampling; probability sampling techniques consist of simple random, stratified and cluster sampling.

This study follows non-probability sampling method and the technique employed to draw the final sample from the sample frame is the judgment sampling. Judgment samples are

¹ Standard Release Procedure- It is a procedure in which a film is first released through movie theaters (theatrical window), then, after approximately 16 and one-half weeks, it is released to VHS and VOD services (entering its video window). After an additional number of months, it is usually released to Pay TV, and approximately two years after its theatrical release date, it is made available for free-to-air TV.

often called the purposive samples: the sample elements are hand-picked because it is expected that they can serve the research purpose. Most typically, the sample elements are selected because it is believed that they are representative of the population of interest (Churchill, 2010).

3.2.4 Sample Size

A sample of five films is selected from the sampling frame. The selection consists of *Paan Singh Tomar* by Tigmanshu Dhulia (2011), *Mary Kom* by Omung Kumar (2014), *Rang Rasiya* by Ketan Mehta (2014), *Manjhi: The Mountain Man* by Ketan Mehta (2015), and *Aligarh* by Hansal Mehta (2016). As mentioned, the technique adopted is judgment sampling, the selection of the final sample is done keeping in mind the following points:

All these films are unique in their own way. *Paan Singh Tomar* celebrates the life of an Indian soldier and athlete who turns into a bagee (rebel) because of certain unfavorable conditions and gets killed in a police encounter. *Mary Kom* is a story of a female boxer who is a five times world boxing champion. Apart from this, the film holds special significance as a biopic because female biopics are quite rare in Bollywood, despite the recent resurgence and newfound popularity of the genre. *Rang Rasiya* is a biopic of painter Raja Ravi Varma and the film waited for its release till 2014 after getting completed in 2008. This film faced a lot of criticism due to censorship and legal issues which are important aspects of biopic filmmaking and filmmakers often find themselves involved in such controversies while projecting a real life on screen. *Manjhi* is a film that goes beyond that convention of biopic filmmaking which has a tradition of celebrating heroes. It narrates extraordinary story of an ordinary man, Dashrath Manjhi, who with his determination, carved a road out of a mountain by working with a chisel for 22 years continuously. Finally, the film *Aligarh* is quite remarkable in the way it deals with the life of Prof. Shrinivas Ramchandra Siras who was a

homosexual. The film stood up for LGBT rights in India and received a mixed public reception.

Biopics and its sub-categories like sports biopics, female biopics, artists' biopics, biopics of the common people etc. have been taken up within the sample so as to encompass not only the different types but also to impart a representative nature to the sample. Thus, the selection is made representative of the different types of biopics which are being produced.

As the study considers the shot-length data which has identified more than 8000 data points, the sample consists of only five biopics. An in-depth stylistic study is carried out and a larger sample in this case can be too laborious to handle which can affect the quality of analysis and a smaller sample would have been inadequate to provide a clear picture of the desired patterns in style and structure.

3.3 Identification of Components to be Measured

Once the final sample of five films is drawn from the selected sample frame, the components or the variables for which data has been recorded or extracted, are identified after a thorough literature review.

As discussed in chapter 2, film style consists of the following elements; Mise-en-scene, Cinematography, Editing and Sound (Bordwell, 2008). For style analysis, the study takes into consideration four components of cinematography; shot duration, shot scale, shot angle and camera movement. These variables are chosen for the study because different studies conducted by researchers such as Barry Salt, Nick Redfern, Jonathan Oliver and others have considered these components significant in determining the style of a film (Oliver 2015, Redfern: 2010, 2013, Salt 1974).

3.4 Data Extraction and Recording Procedure

The data set consists of shot length data and it is extracted manually (Salt 1974) using VLC Media Player. Each film was run in the software and relevant details pertaining to every single shot was recorded which consists of shot duration, shot scale, shot angle and camera movement.

Shot duration was recorded in seconds, shot scale, angle and camera movement were put into different categories and thus recorded accordingly. Shot scale is coded into seven categories: VLS (Very long shot) =1, LS (Long shot) =2, MLS (Medium long shot) =3, MS (Medium shot) =4, MCU (Medium close-up) =5, CU (Close Up)= 6 and BCU (Big close-up) =7 (Oliver, 2015); shot angles are coded into three categories: H (High angle) =1, L (Low angle) =2, N (neutral or eye level)= 3 (Oliver, 2015) and camera movement is coded into four categories; Static shots=1, Simple moving shots= 2 (this includes pan track, tilt, zooms, etc.), Complex moving shots=3 (this includes combinations of simple movements) and Handheld or HH (shaky shots) =4 (Bordwell and Thomson, 2008).

Shot duration is thus a continuous dependent variable and shot scale, shot angle and camera movement are the three independent categorical variables. Data recording is done according to the following criteria:

- The shots are recorded from the point where the narrative action starts, however, the credit titles in the beginning and end of the film are not considered if they do not consist of any narrative action. (Oliver, 2015, p. 5)
- While recording the duration of the shots, the shots which are of duration less than one second, are considered to be of zero seconds, as unit less than a second is not possible to record manually.

- In case of shots which are not static, sometimes with the movement of camera, the scale or the angle changes. This makes it difficult to assign one particular category of scale or angle to a single shot and therefore in such cases, either the first frame/ keyframe of the shot is considered or the overall impact of the shot is considered as per the discretion of the researcher.

When extracted for all the five films, the total number of shots identified for the study are 8293; of which *Aligarh* consists of 565 shots, *Mary Kom* has 2630 shots, *Paan Singh Tomar* has 1774 shots, *Rang Rasiya* and *Manjhi* consist of 1790 and 1534 shots respectively.

3.5 Arrangement of Data for Analysis

The extracted data is arranged at two levels for the sake of an organized and detailed analysis. At the first level, all shots of all the five films are arranged according to the running time and the analysis is done to identify a pattern of use of these variables and the relationship among the variables of each film across the sample.

The data in the first level, which is arranged according to the running time, is also split into four acts (Cutting et. al., 2011) of approximately 30 minutes each (four quarters) to see the patterns of similarities in the act style across the five films. It is dealt with in more detail in Chapter 4, Results and Discussion-I.

For the second level of analysis, the entire data is arranged into eight Narrative components or elements of life; (also referred to as NC) which are given by George F. Custen (1992, p. 67-80).

Custen observes that, “just as the patterned use of one kind of formal element- title cards- characterizes the biopic as a unique film genre, the way the narrative of a life is constructed is also characteristic”. He has enlisted the following Narrative components or the Elements of life: *Family, Close friends and guides, Public Reception of Talent, Innovation, The price*

of being different: Retribution, Fortune or Misfortune, Historical Era depicted, Character Demographics and Ethnicity, Social Class. These are further elaborated in Chapter 5, Results and Discussion-II.

Most of the components are adopted from Custen's (1992) work but since the context of study is Indian, a slight variation in the order has been made and after observing the sample, an addition of two new components, of *Struggle and Success*, has been done as per the requirement of the present study. Also, the components of *Innovation* and *Fortune or Misfortune* are not considered as sequences comprising of these elements are missing in some of the films of the sample.

All the sequences of a film which belong to a particular component are put together and similarly data for all the films of a specific narrative element is extracted for second level of analysis which is discussed in Chapter 5.

3.6. Statistical Style Analysis

The present study is a style analysis of biopics using statistical techniques and the results are computed using the software SPSS 21. The techniques used for the analysis are discussed below.

3.6.1 Descriptive Statistics

Data is explored through descriptive statistics which is a basic technique that provide rich insights into the data and lay the foundation for more advanced analysis. (as quoted in Malhotra and Dash, 2011, p. 437). Descriptive statistics is used to describe the basic features of the data in a study and provides simple summaries about the sample and the measures. Together with simple graphic analysis, it forms the basis of virtually every quantitative analysis of data. This is a univariate analysis which involves the examination across cases of one variable at a time. There are three major characteristics of a single

variable that we tend to look at: the central tendency, the dispersion and the distribution.

(From: Social Research Methods)

3.6.1.1 Descriptive Statistical Measures

These measures are used to describe the characteristics of the sample in totality. (Koul, 2012, 316). The most commonly used statistics associated with frequencies are measures of central tendency (mean, mode and median), measures of variability (range, interquartile range, variance, standard deviation and coefficient of variation) and measures of shape (skewness and kurtosis). These statistical techniques have been discussed below in detail, after referring the works of Churchill (2010), Malhotra and Dash, (2011), Koul (2012) and Field (2013).

3.6.1.2 Measures of Central Tendency

They tend to describe the center of the distribution and for the present study median and mean are considered.

- 1) **Median-** It is the middle value of the sample, when the data is arranged in ascending or descending order; it gives the value above which half of the values fall and below which the other half falls (Koul, 2012, p. 452). Calculation of median gives the shot duration above and below which the shots are distributed in different films and narrative components. Similarly, calculation of other percentiles like 25th and 75th give the values of shot duration below which 25% and 75% of the total shots lie. “Percentiles are also a useful tool. They measure the number of variables at regular intervals of a text. For example, at every five percent, count the number of variables in the film. This reveals whether the variables are evenly distributed throughout the film, or concentrated in a particular part of it.” (Elsseaser and Buckland, 2002, p. 110)

- 2) **Mean-** Mean of a distribution is commonly understood as the arithmetic average. Calculation of mean values in this study gives the average shot duration for each film. It also calculates the average time devoted to each narrative component by different films, and thus gives a very significant basis for comparison. According to Barry Salt “such a measure provides strict comparability between films” (2009, p. 160).

3.6.1.3 Measures of Variability

It is a statistic that indicates the distribution’s dispersion. Dispersion refers to the spread of the values around the central tendency. The common measures of dispersion or variability which are calculated on interval or ratio data, include range, and standard deviation.

Range- It is simply the difference between the largest and the smallest value in the sample. In this study, the range is calculated and compared for the continuous variable of different films. As the range measures the spread of the data, it gives the range of shot durations within which all the shots lie

Standard Deviation- It is a quantity expressing by how much the members of a group differ from the mean value of the group. It is a measure of the dispersion of a set of data from its mean. If the data points are farther from the mean, there is higher deviation within the data set. It is a more accurate and detailed estimate of dispersion because an outlier can greatly exaggerate the range. Since it measures the distribution of values around the mean, a low standard deviation means values occur close to the mean (there is little deviation). Conversely, a high standard deviation means that the values are dispersed a long way from the mean. (Oliver, 2015, p. 14)

3.6.1.4 Measures of Shape

In addition to measures of variability, measures of shape are also useful in understanding the nature of the distribution. The shape of a distribution is assessed by examining skewness.

Skewness- A characteristic of a distribution that assesses its symmetry about the mean. Distributions can be either symmetric or skewed. In a symmetric distribution, the values on either side of the centre of the distribution are the same, and the mean, mode and median are equal. The positive and corresponding negative deviations from the mean are also equal. In a skewed distribution, the positive and negative deviations from the mean are unequal. Skewness is the tendency of the deviations from the mean to be larger in one direction than in the other. A large skew indicates that the majority of values are positioned on one side or other of the mean. If a skew is a negative figure then the majority of the values are found above the mean and if the skew is a positive figure then the majority of the values are found below the mean. (Oliver, 2015, p. 14-15)

Skewness of different groups is compared to observe if there exists a desired pattern in terms of similarity.

3.6.1.5 Frequency Distributions

Frequency distributions are examined with an objective to obtain a count of the number of responses associated with different values of a variable. The relative occurrence, or frequency, of different values of the variable is then expressed in percentages. A frequency distribution of a variable produces a table of frequency counts, percentages and cumulative percentages for all the values associated with that variable. (Malhotra and Dash, 2011). Frequency distributions helps the researcher to explore questions like; which category of shot scale has the maximum usage in a film, use of which category of shot angle or camera

movement dominates the entire film, etc. and these results are then compared with the results of other films in the sample.

In the second level of analysis, while comparing the different narrative components frequency distribution is used to explore questions like what percentage of total shots a film devotes to different components and the results are compared across the sample of five films. It also describes which shot scale category, angle and camera movement is used the most for one component by different films.

3.6.2 Multiple Regression

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. It develops a mathematical relationship between two or more independent variables and an interval-scaled dependent variables. It is a way of predicting an outcome variable (dependent) from predictor variables (independent).

Regression analysis mathematically sorts out which of the variables does indeed have an impact. It answers the questions: Which factors matter most? Which can we ignore? And how do those factors interact with each other? Regression analysis is widely used for prediction and forecasting, but it can also be used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships and it is also used to infer causal relationships between the independent and dependent variables. (Field, 2013)

Multiple Regression involves a single dependent variable and two or more independent variables. This study studies the relationship between the shot duration (dependent variable) and shot Scale, shot Angle and camera movement (independent variables) using the technique of multiple linear regression.

3.7 References

1. Bordwell, D., and Thompson, K. (2008). *Film art: An introduction* (Vol. 8). New York: McGraw-Hill.
2. Borg, W. R., and Gall, M. (1989). *Education research: An Introduction*. New York: Longman.
3. Churchill, Gilbert A., and Dawn Iacobucci. (2010) *Marketing Research: Methodological Foundations*. New York: Dryden Press.
4. Custen, George Frederick. (1992). *Bio/pics: How Hollywood constructed public history*. Rutgers University Press.
5. Cutting, James E., Kaitlin L. Brunick, and Jordan E. DeLong. (2011). How act structure sculpts shot lengths and shot transitions in Hollywood film. *Projections* 5, no.1, 1-16.
6. Cutting, James E., Kaitlin L. Brunick, Jordan E. DeLong, Catalina Iricinschi, and Ayse Candan. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception* 2, no. 6, 569-576
7. Elsaesser Thomas and Warren, Buckland. (2002). *Studying Contemporary American Film: A Guide to Movie Analysis*. London: Arnold.
8. Field, Andy. *Discovering statistics using IBM SPSS statistics*. Sage, 2013.
9. Glass, G. V., and Hopkins, K. D. (1984). Inferences about the difference between means. In *Statistical Methods in Education and Psychology*. Prentice-Hall, Englewood Cliffs, NJ, 230-232.

10. Ji, S., and Waterman, D. (2010). *Production technology and trends in movie content: An empirical study*. Working Paper, Department of Telecommunications, Indiana University, Bloomington, IN. Retrieved from <http://www.indiana.edu>.
11. Koul, Lokesh. (2012). *Methodology of Educational Research*. Vikas Publishing House Pvt Ltd.
12. Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. Longman/Addison Wesley Longman.
13. Malhotra, Naresh K., Dash, Satyabhushan (2011) *Marketing research: An applied orientation*. Pearson Education India.
14. Olliver, J. (2015). Poetics of early Hammer horror films: A statistical style analysis. *New Review of Film and Television Studies*, 13 no 4, 354-384.
15. Redfern, Nick. (2010). Statistical analysis of shot types in the films of Alfred Hitchcock. *Citeseer*.
16. Redfern, Nick. (2013). Film style and narration in Rashomon. *Journal of Japanese and Korean Cinema* 5, no. 1-2, 21-36.
17. Social Research Methods: Retrieved from:
<https://www.socialresearchmethods.net/kb/statdesc.php>
18. Salt, Barry. (1974). Statistical style analysis of motion pictures. *Film Quarterly* 28, no. 1, 13-22
19. Salt, B. (2009). *Film Style and Technology: History and Analysis (Vol. 3rd)*. Totton, Hampshire, UK: Starword.