International Journal of Contemporary Issues in Social Sciences
Volume 3, Issue 2, 2024 ISSN(P):2959-3808 | 2959-2461

# STATISTICAL MODELLING OF T-20 CRICKET AMONG CANCER PATIENTS

**Atta Ur Rehman[1], Qamruz Zaman[2], Danish Wasim[*3], Sofia Mansoor[4]**

[1,2,4]Department of Statistics, University of Peshawar, Pakistan;
[*3]Department of Microbiology, Abasyn University Peshawar, Pakistan

[1]attabaqi11@gmail.com; [2]qamruzzaman@uop.edu.pk; [*3]danishwasim.std@icp.edu.pk ;
[4]sofia@uop.edu.pk

Corresponding Author: *

**Received:** 10 April, 2024 **Revised:** 10 May, 2024 **Accepted:** 28 May, 2024 **Published:** 10 June, 2024

## ABSTRACT

The presence of these variables increases the chance that cancer patients will enjoy playing or watching cricket, even though other factors like age, education level, cancer stage, favorite match, and the extent to which they watched cricket before receiving a diagnosis or attended matches in person cannot be considered significant at the 5% level of significance. The goal of the current Peshawar study was to assess the significant interactions between a variety of variables, including sex and age. Its two goals are to ascertain the correlation between different components as well as investigate the constituents that add to cricket's allure. Primary data were collected through a well-designed questionnaire from the cancer patients in distract Peshawar government hospitals. Logistic regressions were used for examining the factors responsible for the popularity of cricket. Chi square and odds ratios were used to determine the association between different factors.

**Keywords**: Cancer patients, T-20 cricket, Popularity, Stage of cancer, Hospitals.

## INTRODUCTION

Cancer is a devastating disease that affects millions of people worldwide. Cancer statistics for 2020 reveal the alarming incidence and mortality rates associated with various types of cancer. The paper provides valuable insights into the prevalence and impact of cancer, helping healthcare professionals, policymakers, and the general public.

Understand the current state of the disease.

Bray, Ferlay, Soerjomataram, Siegel, Torre, and Jemal (2018) conducted a global analysis of cancer statistics in 2018. Their research highlights the variations in cancer burden across different regions and countries. By estimating cancer incidence and mortality rates for various types of cancer, the study emphasizes the need for international collaboration in addressing the global impact of cancer.

One of the critical aspects of cancer care is addressing the emotional well-being of patients. The study by Carlson, Waller, Groff, Giese-Davis, Bultz, and Hagen (2013) focuses on online screening for distress among newly diagnosed cancer patients. Their research demonstrates that online screening effectively identifies patients experiencing emotional distress, thus enabling early intervention and support. This finding underscores the potential of technology based screening tools in improving the psychosocial care provided to individuals facing a cancer diagnosis.

The popularity of sports games among cancer patients has been recognized for its positive impact on physical and psychological well-being. Engaging in sports activities can provide significant benefits, including improved physical fitness, enhanced quality of life, emotional well-being, and social support.

A meta-analysis of 34 randomized controlled trials (RCTs) by Buffart et al. (2018) examined the effects of exercise on quality of life and physical function in cancer patients. The analysis found that exercise interventions, which often include sports games, had a positive impact on various aspects of quality of life, including physical, social, and emotional well-being. Participating in sports games allows cancer patients to engage in physical activities tailored to their abilities and needs. It helps them maintain or improve physical fitness, which is crucial for managing

treatment side effects and promoting overall well-being. Research by
Wurz et al. (2019) showed that exercise interventions, including sports games, can lead to improvements in physical functioning, aerobic fitness, and muscle strength among cancer patients.
In addition to physical benefits, sports games offer a positive distraction from the challenges of cancer treatment. They provide an opportunity for patients to focus on enjoyable activities, interact with others, and experience moments of joy and accomplishment. A study by Bruun et al. (2014) highlighted the positive psychological impact of sports games, indicating reduced anxiety, depression, and fatigue among cancer patients who engaged in physical activity interventions.
Furthermore, sports games foster a sense of camaraderie and social support among cancer patients. Engaging in these activities provides opportunities for social interaction, creating a supportive community where patients can connect with others facing similar challenges. A study by Stevinson et al. (2009) demonstrated the social benefits of exercise interventions, including sports games, for cancer patients, with participants reporting enhanced social support and improved mood.
Recognizing the benefits, healthcare providers and cancer support organizations have increasingly incorporated sports games into comprehensive cancer care programs. These programs aim to improve the well-being and quality of life of cancer patients by promoting physical activity and providing opportunities for social engagement and emotional support.

### 1.1 Popularity and Benefits of Cricket in Cancer Patients

Cricket, a popular sport worldwide, has shown promise in benefiting cancer patients and providing physical, emotional, and social support. Engaging in cricket-related activities can have a positive impact on the well-being of cancer patients, contributing to their overall quality of life during and after treatment.
Physical benefits: Participating in cricket activities helps cancer patients improve their physical fitness, endurance, and coordination. It allows them to engage in structured physical exercise, which can alleviate treatment side effects and improve overall physical well-being. Research by Fong et al. (2019) demonstrated the positive effects of structured cricket exercise programs on physical function and quality of life in cancer survivors.
Emotional well-being: Cricket can serve as a source of emotional support and provide a positive distraction from the challenges of cancer treatment. Playing cricket and being part of a team can boost self-esteem, enhance mood, and reduce psychological distress. A study by Galva˜o et al. (2019) explored the psychological benefits of cricket among cancer survivors, highlighting improvements in mood, self-confidence, and body image.
Social support: Cricket offers cancer patients an opportunity to connect with others facing similar challenges, fostering a sense of camaraderie and social support. Being part of a cricket team creates a supportive environment where patients can share experiences, build friendships, and feel a sense of belonging. The social aspects of cricket can help reduce feelings of isolation and enhance overall well-being. Research by Kirkman et al. (2020) emphasized the importance of social support and camaraderie in cricket-based interventions for cancer survivors.
Implementing cricket programs: Healthcare providers, cancer support organizations, and cricket associations can collaborate to develop structured cricket programs specifically tailored for cancer patients. These programs should consider the individual needs and abilities of patients, providing appropriate modifications and accommodations. Cricket programs can be integrated into comprehensive cancer care, promoting physical activity, emotional well-being, and social connectedness.

## 2. Methodology

This chapter describes the study universe, sample design, data collection technique, and statistical tools that will be used to analyse and collect data for the research study's set proposals.
The approaches are described in the following segments.

### 2.1 The study universe

This research was conducted in the District Peshawar, KPK.
The population consist of all patients especially cancer patients, since the size of the population is unknown, therefore district Peshawar hospitals was consider for data collection.

**2.2 Sampling Design**

Cluster sampling is a common method used in survey research to efficiently collect data from large and diverse populations. For this research study, a cluster sampling technique will be employed in Peshawar (target population) to ensure adequate representation of the target population.

**2.3 Collection of Data**

The current research study used primary data in a well-designed structure questionnaires from different cancer hospitals located in District Peshawar KPK. The inquiry form was designed to cover all relevant information on cancer patients and their interest in T-20 cricket, which is required to meet the research study's objectives.

The majority of the questions in the developed questionnaire are binary in nature, i.e. YES and NO. The target population of this study is all cancer patients aged between 15 to 40.

The Cochran formula for calculating the appropriate sample size has been used for this study. The sample size calculated was 385 with a 95 % confidence interval and with 5% margin of error. However, we take a sample size of 500 participants because this research study on the popularity of the cricket game among cancer patients aims to provide robust and representative findings. The large sample size allows for a more comprehensive analysis, enhancing the generalizability of the study's results.

**2.4 Statistical Analysis**

This section of the research study includes the statistical tools used to analyze the acquired data and get research-based conclusions in order to conclude the research's major purpose. The Statistical Package for Social Sciences (SPSS) V.21 was used to analyze the gathered data. The

section on analysis is broken into three parts. The frequency distribution and descriptive metrics of the researched features are included in the first section. The second section comprises the odds ratio and chi-square statistics, which demonstrated the relationship of various cancer patient variables. Similarly, the odds ratio was calculated to evaluate the impact of risk factors on cancer status. Furthermore, the findings of logistic regression are shown in the third section of the analysis. The logistic regression analysis was used to identify cancer risk variables. The following section talks over all of these statistical techniques

**.2.4.1 Odds ratio and Chi square test**

The chi-square test, also known as Pearson's chi-square test, was used to assess the link between cancer and all of the factors considered in this study. The hypothesis (i) defined in section (1.4) was tested using this test. This test assumed no major correlation with the specified factors under the null hypothesis, and otherwise under the alternative hypothesis. Equation defines the general formula for chi-square tests for testing the association. (2.1)

$$\chi^2 = \sum\sum \frac{(O_{JK} - E_{JK})^2}{E_{JK}} \quad (2.1)$$

Where $o_{jk}$ and $E_{jk}$ represent the observed and expected frequencies, which correspond to the $j^{th}$ and $K^{th}$ columns of the contingency table, respectively; and r and c denote the number of rows and columns, respectively.

The significance of the chi square test results for examining the association of cancer with the specified factors was determined by taking the 5% and 1% level of significance into account. If the P-value for a chi square test was less than or equal to, the test was considered significant; otherwise, it was considered inconsequential..

In order to compute the odds ratio, let us consider the following 2 × 2 contingency table.

| Group | Success | Failure | Total |
|---|---|---|---|
| $K_1$ | A | B | A + B |
| $K_2$ | C | D | C + D |
| Total | A + C | B + D | n= A+ B + C + D |

The odds ratio, denoted by $\psi$ is defined as:

$$\psi = \frac{AD}{BC} = \frac{P1/(1-P1)}{P2/(1-P2)}$$

,P1 and P2 represent the success probabilities of groups 1 and 2, respectively. The odds ratio value is compared to (1) to determine the odd odds for groups 1 and 2.If the odd ratio value is greater than one, the odds of group-1 are greater than those of group-2. If the odd ratio value is less than one, the odds of group-2 are greater than those of group-1.

**2.4.2 Logistic Regression**

Logistic regression models are used when a model's class or event (response variable) has a binary possibility, such as win or lose, pass or fail, healthy or sick, etc. In this case, the answer variable was assigned a value of 1 or 0 depending on whether a

certain feature was present or absent. When the response variable is binary, logistic regression is the best model for analysing the impact of various risk variables.

Singoyl et al. (2016) did a multivariate logistic regression analysis of independent cancer risk variables.

Equation (3.2) defines multiple linear regression models as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon \quad (2.2)$$

$$= \beta_0 + \sum \beta i x i$$

In the following equation, Y is the dependent/response variable, X1, X2,....., Xk are independent variables, β0 is the intercept term, beta1, β2,...., βk are the K-independent variable regression coefficients, and is the residual term with a normal distribution, zero mean, and constant variance.

Taking expectation on both sides of the equation (2.2), it can be reduced to:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \quad (2.3)$$

The response variable in equation (2.3) is binary, with values 1 for yes and 0 for no.

The anticipated value of Y for a binary response variable is 1, and it may be represented in terms of probability as:

E(Y) = 1,P(Y=0) + P(Y=1)= P(Y=1)

Given that P(X) represents the probability of a K-independent variable, the variance of the response variable can be written as follows:

$$var(Y) = E(Y^2) - (E(Y))^2 = P(x).P[1 - p(x)] \quad (2.4)$$

Where, $E(Y^2) = 1^2 P(x) + 0^2 P[1 - p(x)]$

By putting P(x) = P,then the variance of Y is expressed as: Var(Y) = P[1-P]

Now taking equation (2.3), it can be expressed as:

$$E(Y) = P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

In terms of logit,it can be written as:

$$Logit(p) = In(p/1 - p) \quad (2.5)$$

Then equation (2.3) can be written as:

$$In(P/1 - P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_K \quad (2.6)$$

Equation (2.6) is known as the logistic regression model, and it can also be stated as shown in equation (2.7), which is known as the linear probability model.

$$P = \frac{exp[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k]}{1 + exp[\beta_0 + \beta_1 X_1 + \beta_2 X_2]} \quad (2.7)$$

$$= \exp[\beta_0 + \sum_{i=1}^{k} \beta_i X_I] / 1 + \exp[\beta_0 + \sum_{i=1}^{k} \beta_i X_I]$$

Equation 2.7 can be written for a specific independent variable (x) as follows

$$P = \exp(\beta_0 + \beta_x)/1 + exp(\beta_0 + \beta_x)$$

Changing the independent variable by one unit increases the probability of Y by eβ times.

2.4.3 Conclusion for the model of Logistic regression

The inferences from the model of logistic regression will be produced simply by using a straightforward linear regression model. Wald derived the typical results of most evaluators for logistical regression in 1943, with the model length believed to be appropriate. So, for the bulky pattern, the 100% confidence c in the programming language for exponential version coefficients " will be conveyed as:

$$b \pm Z_{\alpha/2} SE(b) \quad (2.8)$$

Where "b" is the estimated value of β.

Let us recall the logistic regression model with unknown regression parameters 1,2,....k, and assume that our goal is to compare the hypothesis i = 0 to the alternative i = 0. We utilised M1 for the fitted logistic regression model and M2 for the modest model. Assuming that (L1) and (L2) are the likelihood functions for M1 and M2, respectively, the test statistic represented by (L1 - L2) with 'K' degree of freedom is used to test this hypothesis (Wilks, 1938).

### 2.4.4 Fitted Model diagnostics

Model parameters are predicted during the analysis procedure. Trying out the appropriateness of the general model is a critical situation of evaluation, and identifying how well the equipped model accurately portrayed the found facts is a key component of the modelling technique. When one version contains specific linear regression but their relationship is not linear, the parameters predicted and conclusions drawn may be invalid, thus it is vital to determine whether the information helps the version in question or not. The diagnostic technique determines whether the geared-up version is adequate. The following are a few of the most widely utilised diagnostic procedures for determining the best and most unique version that was used.

**2.4.5 Index-plot for leverage values**

In regression, leverage values are defined as the across aspect of matrix projection as follows:

$H = W1 / 2X (X1WX)-1XW1/2$ (2.9)

Where "W" represents the load lattice employed in this model, and (X) represents the statement's framework. In an influence esteem, howdy refers to the departure of the observation from the focal point of the information, and these attributes provide a range to the perceptions, allowing a given value to be exceptional or not. Values can also be used to discern the persuasive influence. If we removed the compelling perception from the characteristics. Those perceptions are now virtually a huge number of times, there are long good ways from amazing perceptions and referred to powerful perception. The correlation is carried by using a "influence plot," which is obtained by plotting each impression with their corresponding factors.

**2.4.6 Index plot of residuals**

"Usually, a plot that shows the difference between expected and actual values is created to test how well the model fits the data." When the residual plot follows a straight line pattern, the model is most accurate. It is difficult to choose the optimal model if the residuals plot does not display a straight line pattern." I hope this was helpful! Please let me know if you have any more inquiries. The difference between observed and projected worth is referred to as residuals. It can provide insights and residuals based on detailed data going to fit of adequacy. In a logistic regression framework, the residuals of "n" perceptions with fitted characteristics y I = np'i compared with the react factor "y" for it view can be represented as I = yi-y. Pearson residuals are defined as; when these residuals can be separated by a standard error in the reaction variable yi.

$$Xi = \frac{E}{\sqrt{\text{nPi}(1-\text{PI})}}$$

Their Pearson Chi-square trial sum of squares by data = Pxi, which can also be employed to measure the very well of fit test for logistic regression..

**3. Results and Discussion**

**Table 3.1: Association of risk-factors with enjoy watching or playing cricket**

| Factors | Category | Watching Cricket (% age) | | Odd ratio | Chi-square | p-value |
|---|---|---|---|---|---|---|
| | | YES | NO | | | |
| **Gender** | Male | 273 (54.6) | 72 (14.4) | 6.626 | 6.209 | .013 |
| | Female | 137 (27.4) | 18 (3.6) | | | |
| **Age** | 15-20 years | 178 (35.6) | 43 (8.6) | 1.047 | 1.038 | 0 .904 |
| | 21-25 years | 108 (21.6) | 20 (4) | | | |
| | 26-30 years | 44 (8.8) | 9 (1.8) | | | |
| | 31-35 years | 40 (8) | 10 (2) | | | |
| | 36-40 years | 40 (8) | 8 (1.6) | | | |
| **Education level** | Illiterate | 45 (81.81) | 10 (18.18) | 0.001 | 0.001 | 0.970 |
| | Literate | 365 (82.022) | 80 (17.97) | | | |
| **Marital status** | Married | 153(30.1) | 48 (9.6) | 10.785 | 12.407 | 0.002 |
| | Unmarried | 257 (51.4) | 42 (8.4) | | | |
| **Stage of cancer** | 1st | 254 (50.8) | 52 (10.4) | 0.663 | 0.674 | 0.879 |
| | 2nd | 74 (14.8) | 18 (3.6) | | | |
| | 3rd | 54 (10.8) | 14 (2.4) | | | |
| | Last | 28 (5.6) | 6 (1.2) | | | |
| **Favorite game** | CRICKET | 178 | 43 | 1.047 | 1.038 | 0.904 |
| | FOOD BALL | 108 | 20 | | | |
| | BADMINTON | 44 | 9 | | | |
| | 31-35 years | 40 | 10 | | | |
| | 36-40 years | 40 | 8 | | | |

Table 3.1 shows the link between enjoying watching cricket and other risk variables using the chi square test and the odds ratio test. Out of 500 cancer patients, 273 (54.6%) are male and the remaining 137 (27.4%) are female, while in the control group, males account for 72 (14.4%) and females account for 18(3.6%).It shows that the percentage of males who appreciate cricket is higher than the percentage of girls who watch cricket. The chi square with 6.209 and OR = 6.626 indicating that male are 6.626 times more at risk to be infected as compare to female. The p-value = 0.013 It means that there is a significant association of enjoy watching or playing cricket with gender.

The relationship between the age and enjoy watching or plying cricket are shown in Table 4.10 shows the link between enjoying watching cricket and other risk variables using the chi square test and odds ratio test. Out of 500 cancer patients, 273 (54.6%) are male and the remainder 137 (27.4%) are female, while in the control group, males are 72 (14.4%) and females are 18 (3.6%).It suggests that the percentage of males who appreciate cricket is higher than the percentage of ladies who watch cricket.

Table 3.1 shows the link between enjoying watching cricket and other risk variables using the chi square and odds ratio tests. Out of 500 cancer patients, 273 (54.6%) are male and the remainder 137 (27.4%) are female, while in the control group, male are 72 (14.4%) and female are 18 (3.6%).It shows that the proportion of males who appreciate cricket is higher than the proportion of girls who watch cricket.

Table 3.1 also shows the results of the association between the individuals' marital statuses and watching or playing cricket. It is clear that out of 500 patients, 257 (51.4%) are not married and 153 (30.1%) are married, whereas in the control group, 42 (8.4%) are not married and 48 (9.6%) are married. It is obvious that the percentage of patients who are not married is higher than that of married people (OR = 10.785), implying that unmarried people are 10.785 times more likely to watch or play cricket than married people. It also gives chi square of 12.407 with p value 0.002. Overall, marital status was significantly associated with enjoying watching or playing cricket.

Table 3.1 also contains the results of association between stage of cancer of the subjects with enjoy watching or playing cricket. Out of 500 patients 254 (50.8%) having 1st stag, 74 (14.8%) stage 2nd, 54(10.8%) stage 3nd and the last stage have 28 (5.6%).

While 10.4%, 3.6%, 2.4% and 1.2% are not enjoy watching or playing cricket.

The odds ratio is 0.663 which indicate that 0.663 times more patients watching cricket while The chi square value =0.674 with p-value =0.879 it shown that there is no significance association between the enjoy watching or plying cricket with stage of cancer.

Table 3.1 also contains the results of association between stage of cancer of the subjects with enjoy watching or playing cricket. Out of 500 patients 254 (50.8%) having 1st stag, 74 (14.8%) stage 2nd, 54(10.8%) stage 3nd and the last stage have 28 (5.6%).

While 10.4%, 3.6%, 2.4% and 1.2% are not enjoy watching or playing cricket.

The Odds ratio is 0.663 which indicate that 0.663 times more patients watching cricket while the chi square value =0.674 with p-value = 0.879 it shown that there is no significance association between the enjoy watching or plying cricket with stage of cancer.

**Table 3.2: Likelihood ratio and various R square**

| Model | -2likelihood | Cox and Snell square | Nagelkerke R square |
|---|---|---|---|
| 1 | 369.432 | 0.184 | 0.302 |

The fitted model and tables for selected significance risk factors of enjoying watching or playing cricket are as follows. Table 3.2 shows that the final step -2likelihood ratio value is 369.432.

**Table 3.3 Classification table with predictors**

| **Classification Table** | | | | | |
|---|---|---|---|---|---|
| | Observed | | Predicted | | |
| | | | do you enjoy watching or plying cricket | | Percentage Correct |
| | | | No | Yes | |
| Step 1 | do you enjoy watching or plying cricket | No | 28 | 62 | 31.1 |
| | | Yes | 9 | 401 | 97.8 |
| | Overall Percentage | | | | 85 .8 |

Table 3.3 provides the estimated prediction percentage for the fitted model. SPSS provides two stages by default for producing logistic regression. The following table indicates the two options yes and no. The overall percentage of predicated value is 85.8 whereas; the percentage of no is 31.1 and the percentage of yes 97.8. Which means that most of the patient are like to watch cricket.

**Table 3.4: Variable in the best fitted model**

| Variable in the model | *B* | S.E | Df | Sig | Exp (*β*) |
|---|---|---|---|---|---|
| Played or watched cricket before diagnosis. $x_1$ | -.838 | .381 | .028 | .433 | .205 |
| Cricket is gaining popularity worldwide.$x_2$ | -.061 | .361 | .866 | .941 | .464 |
| Ever attended a live cricket match.($x_3$) | 1.410 | .349 | .000 | 4.098 | 2.067 |
| Prefer watching domestic or international cricket matches($x_4$) | -.366 | .373 | .326 | .693 | .334 |
| cricket is a good form of entertainment($x_5$) | -1.519 | .538 | .005 | .219 | .076 |
| Does watching cricket help in reducing your stress level.($x_6$) | 2.776 | .773 | .000 | 0.03 | 3.529 |
| Would you like to see more cricket matches being played in hospitals or care facilites($x_7$) | 1.816 | .356 | .000 | 0.031 | 3.059 |
| Do you think cricket can help in promoting social interactions among patients? ($x_8$) | 2.332 | .883 | .008 | 0.034 | 1.823 |
| Do you think cricket is a good way to relax or distract yourself? ($x_9$) | -.946 | .477 | .047 | .388 | .152 |
| Do you think watching or playing cricket has a positive impact on your mental health.($x_{10}$) | -.485 | .463 | .295 | .616 | .248 |
| Have you ever participated in cricket match as a part of a physical rehabilitation program ($x_{11}$) | -.124 | .289 | .667 | .034 | .501 |
| More opportunities for playing or watching cricket should be offered in health care setting($x_{12}$) | -.472 | .411 | .251 | .624 | .279 |
| Cricket can be used as a tool for socialization and community building among patients($x_{13}$) | -.017 | .306 | .957 | .040 | .540 |
| Watching or playing cricket has any impact on your mood.($x_{14}$) | 0.024 | .319 | .941 | .023 | .522 |
| Would you be interested in participating in a cricket match or tournament with other patients.($x_{15}$) | .419 | .369 | .256 | 1.521 | .737 |
| Cricket is a good sport for promoting teamwork and social interaction. ($x_{16}$) | 0.108 | 0.650 | 0.868 | 1.114 | 0.031 |
| Have you ever participated in a cricket tournament or competition? ($x_{17}$) | 0.075 | 0.305 | 0.806 | 1.078 | .039 |
| Have you ever played cricket yourself($x_{18}$) | -0.810 | 0.341 | 0.018 | 0.044 | 0.228 |
| Cricket is a good form of exercise for patients($x_{19}$) | 0.417 | 0.410 | 0.309 | 0.050 | .060 |
| Cricket can have a positive impact on a patient mental and physical health. ($x_{20}$) | 0.251 | 0.483 | 0.603 | 0.778 | .0302 |
| Constant | 0.533 | 0.331 | 0.108 | 0.023 | |

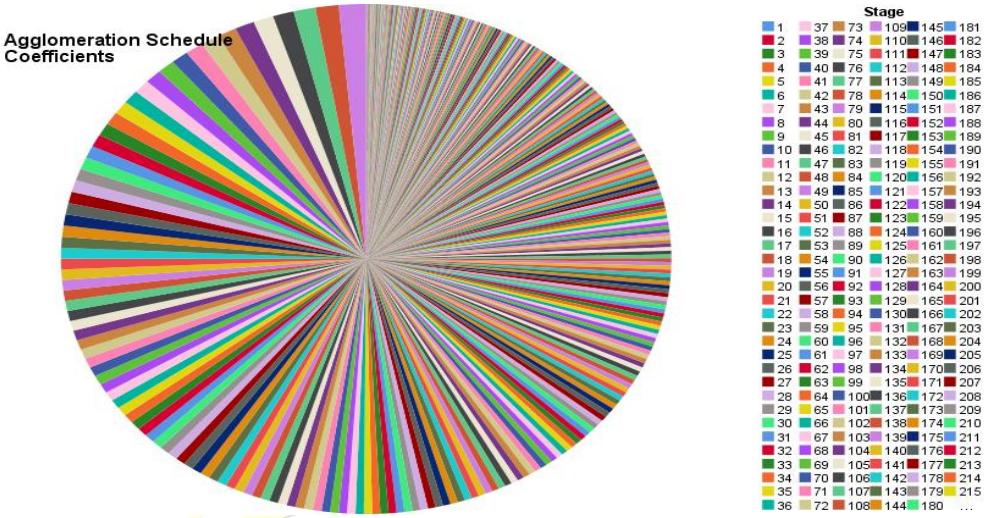
Table 3.4 displays the results of the best fitted model, including the selected components' regression coefficients, standard error, level of significance, odd ratio (Exp(B)), and 95% confidence intervals. From the model it is clear that the factors of "have you ever played or watched cricket before your diagnosis" has

a negatively insignificant (p=0.433) association with enjoy playing or watching cricket with the risk factor is 0.205 indicating negatively insignificant effect on the dependent variable. One unit change in “Cricket is gaining popularity worldwide” causes -0.061 decreases in dependent Variable inserted with standard error of 0.361 and showing the insignificant effect on the dependent variable. Furthermore, a one-unit change in "Have you ever attended a live cricket match" results in a 1.410 increase in "Do you enjoy watching or playing cricket matches" with a standard error of.349, indicating an inconsequential influence on the dependent variable. One unit change in "Do you prepare watching domestic or international cricket matches" results in a -.366 reduction in the dependent variable, with a standard error of 0.373, indicating an insignificant influence on the dependent variable. While a one-unit change in "Do you think cricket is a good form of entertainment" results in a -1.519 reduction in the dependent variable, with a standard error of 0.538 and a negligible effect on the variable. One unit change in "Does watching cricket health in reducing your stress level" results in a 2.776 increase in the dependent variable, with a standard error of.773, indicating a substantial effect on the variable. Furthermore, one unit change in "would you like to see more cricket matches being played in hospitals or care facilities" results in a 1.816 increase in the dependent variable, with a standard error of 0.356, indicating a significant effect on the dependent variable. One unit change in "Do you think cricket can help in promoting social interactions among patients" results in a 2.332 rise in the dependent variable, with a standard error of 0.883, indicating a substantial influence on the dependent variable.

One unit change in "Do you think cricket is a good way to relax or distract yourself" results in a -0.946 drop in the dependent variable, with a standard error of 0.477, indicating an inconsequential influence on the dependent variable. While a one-unit change in "Do you think watching or playing cricket has a positive impact on your mental health" results in a -0.485 drop in the dependent variable, with a standard error of 0.463, indicating that the effect is small. One unit change in "Have you ever participated in a cricket match as part of a physical rehabilitation programmer" results in a -0.124 drop in the dependent variable, with a standard error of 0.289, indicating a negative negligible effect on the dependent variable. Furthermore, a one-unit change in "Do you think more opportunities for playing or watching cricket should be offered in health care setting" results in a -0.472 drop in the dependent variable, with a standard error of 0.411 and a negligible effect on the dependent variable. One unit change in "Do you think watching or playing cricket has any impact on your mood" increases the dependent variable by 0.024, with a standard error of 0.319, indicating a substantial effect on the dependent variable. One unit change in "would you be interested in participating in a cricket match or torment with other patient" results in a 0.419 increase in the dependent variable, with a standard error of 0.369, indicating that the effect on the dependent variable is small. One unit change in "Do you think cricket is a good sport for promoting teamwork or social interaction?" results in a 0.0108 rise in the dependent variable, with a standard error of 0.650, indicating a negligible influence on the dependent variable. One unit change in "Have you participated in a cricket torment or competition" results in a 0.075 increase in the dependent variable, with a standard error of 0.305, indicating a negligible influence on the dependent variable. One unit change in "Have you ever played cricket yourself" causes a -0.810 decrease in the dependent variable added with standard error 0.341, indicating a negative effect on the dependent variable, whereas the other unit change in "Do you think cricket is a good form of exercise for patients" causes a 0.417 increase in the dependent variable added with standard error 0.410, indicating a significant effect on the dependent variable. One unit change in "Do you think cricket can have a positive impact on patient mental and physical health" causes 0.251 increase in dependent variable added with standard error 0.483 and showing the insignificant effect on the dependent variable, while the other unit change in "Have you participated in a cricket torment or competition" causes 0.075 increase in dependent variable added with standard error 0.305 and showing the insignificant effect on the dependent variable.

By using the results, the best fitted subset model for predicting the enjoy watching or plying cricket is express as:

$$logit\ (\hat{p}) = 0.533 - 0.838 - 0.061x_2 + 1.410x_3 - 0.366x_4 - 1.519x_5 + 2.776x_6 + 1.816x_7 + 2.332x_8 - 0.946x_9 - 0.485x_{10} - 0.124x_{11} - 0.472x_{12} - 0.017x_{13} - 0.024x_{14} + 0.419x_{15} + 0.108x_{16} + 0.075x_{17} - 0.810x_{18} + 0.417x_{19} - 0.251x_{20} + £ .$$




b. Ward Linkage

**Clustering:**

**Hospitals wise**

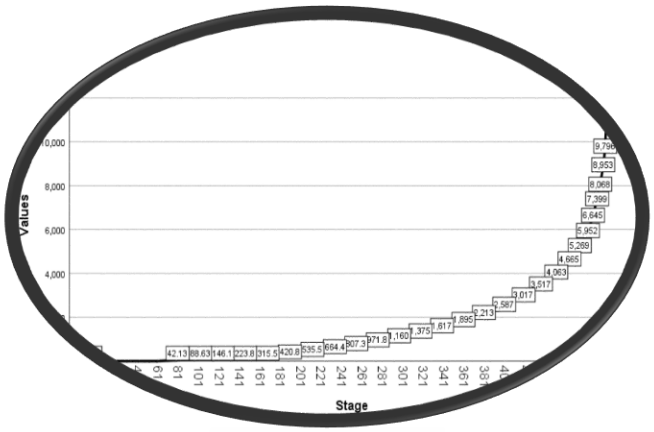
**Case Processing Summary**

**Cases**

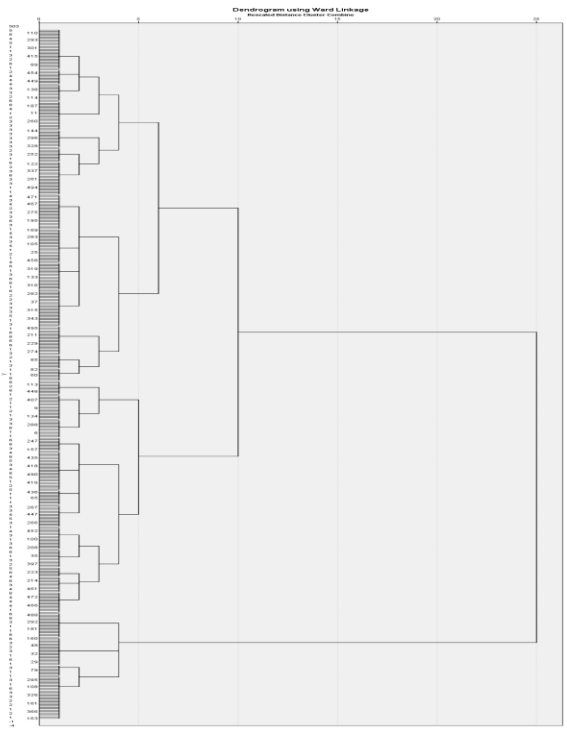
| Valid | | Missing | | Total | |
|---|---|---|---|---|---|
| **N** | **Percent** | **N** | **Percent** | **N** | **Percent** |
| 500 | 100.0 | 0 | .0 | 500 | 100.0 |

a. Squared Euclidean Distance used

The "Case Processing Summary" table indicates that there are a total of 500 cases, with no missing data, thus making all 500 cases valid and complete. This represents a 100% valid case rate, with no instances of missing data, reflecting a comprehensive and complete dataset for analysis. The absence of missing data simplifies data interpretation and enhances the reliability of findings derived from this dataset.




Certainly! The bar graph titled "Agglomeration Schedule Coefficients" shows the values across different stages of an agglomeration process. The horizontal axis represents the stages (numbered from 1 to 15), while the vertical axis represents the corresponding values. Notably, the highest value occurs at stage 1 (approximately 11,477), gradually decreasing as the stages progress. Stage 15 has the lowest value (approximately 2,235). This graph provides insights into how specific coefficients change during the agglomeration process



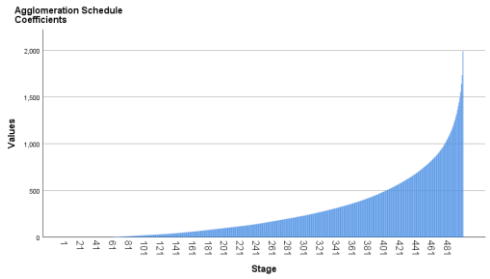

The dendrogram you've generated using Ward Linkage presents a hierarchical structure where clusters are formed by progressively combining smaller clusters to minimize within-cluster variance. Once populated with data points, this dendrogram will reveal distinct clusters labeled for identification. By illustrating the merging process with connecting lines, it visually communicates the underlying hierarchical relationships among data points, emphasizing cluster composition and groupings. To enhance interpretability, incorporating color coding and clear labels will provide a more informative and visually appealing representation of the clustering patterns, aiding in the understanding of complex data structures and relationships within the dataset.

**Area wise clustering:**

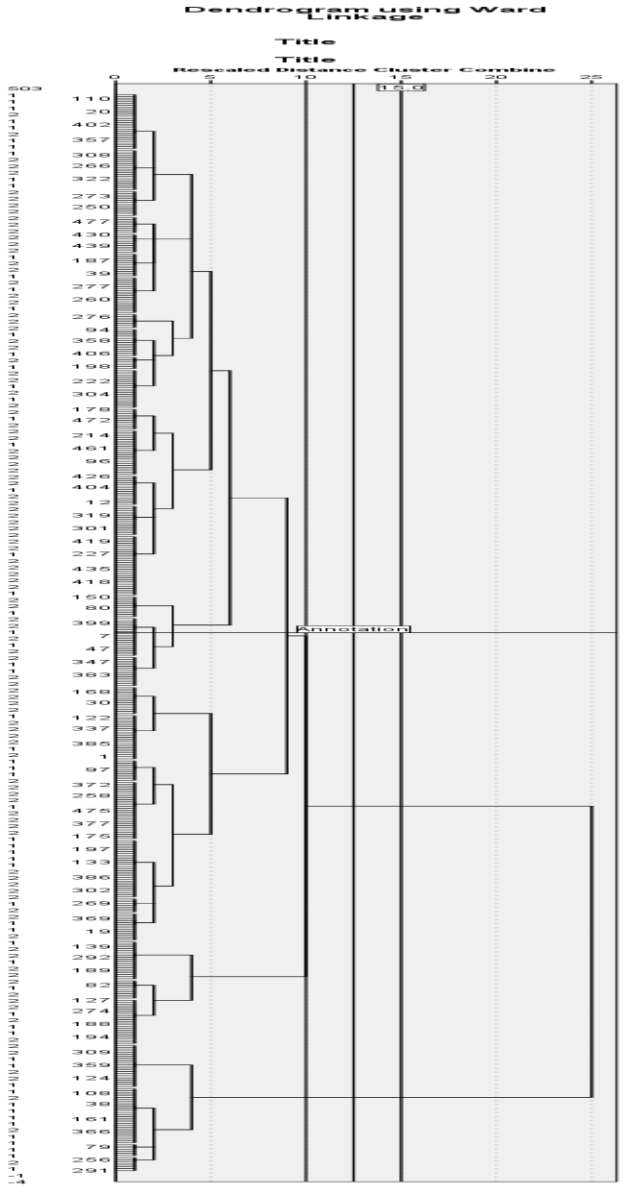
| Case Processing Summary | | | | | |
|---|---|---|---|---|---|
| Cases | | | | | |
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| 500 | 100.0 | 0 | .0 | 500 | 100.0 |

a. City Block Distance used

The case processing summary indicates that there are a total of 500 cases, with all cases being valid and none missing, accounting for 100% of the total cases. This summary suggests that the dataset is complete, with no missing data points to address or impute. The 100% valid cases indicate that all required data fields are populated for each case, ensuring a robust dataset for analysis without the need for data cleaning or remediation due to missing values.




The image depicts a graph titled "Agglomeration Schedule Coefficients," with the x-axis representing stages 1 to 49 and the y-axis showing values from 0 to 2,000. The graph features a blue bar that sharply increases at stage 49, suggesting a significant change in the data at this point. This pattern is typical in hierarchical clustering analyses, where it can indicate the optimal number of clusters by highlighting a substantial increase in dissimilarity between clusters at a specific stage. The graph serves as a critical tool for interpreting data relationships and cluster formation in studies.




.

The image depicts a dendrogram created using Ward Linkage, a method for hierarchical clustering. The x-axis labels are unclear, but the y-axis represents the distance or dissimilarity between clusters. The dendrogram branches out horizontally and vertically, illustrating relationships among different data clusters. While the specific clusters are not discernible, the overall structure highlights how data points group together based on similarity. The background is white with grid lines for better visualization. In summary, this dendrogram visually represents the hierarchical organization of data clusters, emphasizing their proximity or dissimilarity

**Conclusion**

According to the findings, a greater proportion of male cancer patients than female patients said they like to watch cricket. There is a substantial correlation between gender and enjoying cricket, according to the odds ratio and chi-square test. On the other hand, no meaningful correlation was discovered between age and cricket enjoyment. Gender: According to the research, cancer patients' enjoyment of watching cricket may be influenced by their gender. Compared to women, men are more likely to love cricket. Numerous variables, like cultural preferences, social standards, or individual interests, may be to blame for this.

Age: The findings did not reveal a significant correlation between age and enjoying cricket, despite the fact that different age groups experienced differing degrees of enjoyment. This suggests that cancer patients' pleasure of cricket may not be much influenced by their age.

The factor "Have you ever played or watched cricket before your diagnosis" had a negatively insignificant association with enjoying playing or watching cricket. This suggests that prior experience with cricket before a diagnosis did not significantly impact the enjoyment of the sport. 4. The factor "Cricket is gaining popularity worldwide" also had a negatively insignificant association with enjoying playing or watching cricket. This indicates that the perception of cricket's global popularity did not significantly affect the enjoyment of the sport. 5. On the other hand, the factors "Have you ever attended a live cricket match" and "Does watching cricket help in reducing your stress level" showed a positive and significant effect on enjoying playing or watching cricket. This implies that attending live matches and finding stress relief through watching cricket had a significant impact on the enjoyment of the sport.

The combination of the dendrogram diagram showcasing Ward Linkage and the case processing summary table provides a comprehensive view of data analysis and clustering. The dendrogram illustrates hierarchical clustering, offering insights into how clusters form based on similarities between data points. With Ward Linkage, the focus is on minimizing variance within clusters, leading to compact and well-defined groupings. The case processing summary complements this by confirming a complete dataset with no missing values, ensuring the validity and integrity of the data used for clustering. Together, these elements demonstrate a thorough approach to data exploration and analysis, laying a strong foundation for further interpretation and insights.

**References**

Atiq, F., Mauser-Bunschoten, E. P., Eikenboom, J., van Galen, K. P., Meijer, K., de Meris, J., ... & WiN Study Group. (2019). Sports participation and physical activity in patients with von Willebrand disease. *Haemophilia*, *25*(1), 101-108.

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits?. *Medicine & Science in Sports & Exercise*, *33*(6), S379-S399.

Cokkinides, V., Albano, J., Samuels, A., Ward, M., & Thum, J. (2005). American cancer society: Cancer facts and figures. *Atlanta: American Cancer Society*, *2017*.

Francavilla, G., Abrignani, M. G., Braschi, A., Sciacca, R., Francavilla, V. C., Caracciolo, M. M., ... & Braschi, G. (2007). Physical exercise and sport activities in patients with and without coronary heart disease. *Monaldi archives for chest disease*, *68*(2).

Fromme, A., Dreeskamp, K., Pollmann, H., Thorwesten, L., Mooren, F. C., & Völker, K. (2007). Participation in sports and physical activity of haemophilia patients. *Haemophilia*, *13*(3), 323-327.

Idaya Rani, C., & Subbu Lakshmi, M. (2023). Mental fitness: psychological warfare from battlefield to playground. *Psychological studies*, *68*(2), 190-196.

Owen, N., Bauman, A., & Brown, W. (2009). Too much sitting: a novel and important predictor of chronic disease risk?. *British journal of sports medicine*, *43*(2), 81-83.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., ... & Wilmore, J. H. (1995). Physical activity and public health: a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Jama*, *273*(5), 402-407.

Rani, C. I., & Lakshmi, M. S. (2023). Sport and Myth: Reviving Indian Ancient Athletes' Self-Concept in Psychological and Spiritual Dimensions. *Human Arenas*, 1-13.

Redfern, J., & Briffa, T. (2011). Cardiac rehabilitation–moving forward with new models of care. *Physical Therapy Reviews*, *16*(1), 31-38.

Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: a unifying theory. *Clinical psychology review*, *21*(1), 33-61.

Shephard, R. J. (1990). Physical activity and cancer. *International journal of sports medicine*, *11*(06), 413-420.

Smith, E. L., & Gilligan, C. (1991). Physical activity effects on bone metabolism. *Calcified tissue international*, *49*, S50-S54.

Ströhle, A. (2009). Physical activity, exercise, depression and anxiety disorders. *Journal of neural transmission*, *116*, 777-784.

Waardenburg, M. (2006). Active HIV awareness: a study into the meanings of sports as a medium of HIV awareness in a South African township. *Utrecht School of Governance, Utrecht University*