|  |  |  |  |
| --- | --- | --- | --- |
| **Table-1 Frequency Analysis** | | | |
|  | **Variables** | **Frequency** | **Percent** |
| Region | Indian Ocean | 108 | 21.6 |
|  | Mediterranean Sea | 139 | 27.8 |
|  | North Atlantic | 117 | 23.4 |
|  | Pacific Ocean | 136 | 27.2 |
|  | *Total* | *500* | *100* |
| Breeding\_Season | Monsoon | 177 | 35.4 |
|  | Winter | 161 | 32.2 |
|  | Summer | 162 | 32.4 |
|  | *Total* | *500* | *100* |
| Fishing\_Method | Line | 180 | 36 |
|  | Net | 162 | 32.4 |
|  | Trawl | 158 | 31.6 |
|  | *Total* | *500* | *100* |
| Overfishing\_Risk | No | 255 | 51 |
|  | Yes | 245 | 49 |
|  | *Total* | *500* | *100* |
| Species\_Name | Cod | 62 | 12.4 |
|  | Herring | 78 | 15.6 |
|  | Mackerel | 57 | 11.4 |
|  | Salmon | 68 | 13.6 |
|  | Sardine | 70 | 14 |
|  | Shark | 55 | 11 |
|  | Snapper | 52 | 10.4 |
|  | Tuna | 58 | 11.6 |
|  | Trout | 58 | 11.6 |
|  | *Total* | *500* | *100* |
| Water\_Pollution\_Level | High | 167 | 33.4 |
|  | Low | 148 | 29.6 |
|  | Medium | 185 | 37 |
|  | *Total* | *500* | *100* |

Frequency distribution for selected variables is provided in the Table 1. By frequency analysis of variable Region, it appears that a significant portion of observations relate to Mediterranean Sea - 27.8% and Pacific Ocean - 27.2%, then North Atlantic with 23.4% and the Indian Ocean-21.6% consequently. This may show relatively even scattering in the datasets across different regions.

In the case of Breeding\_Season, most of the cases are in Monsoon, comprising about 35.4%, followed by Winter with 32.2% and Summer with 32.4%. This reflects a slight preference for Monsoon as the breeding period among the studied populations.

Among the Fishing\_Method, Line fishing was the most preferred one 36%, followed by Net fishing constituting 32.4%, and Trawl fishing made up 31.6%. This shows that all three types of fishing are almost equally prevalent, with a slight inclination towards Line fishing.

The variable Overfishing\_Risk shows that 51% are not overfishing, whereas 49% of them do. This means nearly half of the species are in critical danger due to overfishing.

The data on Species\_Name indicates that Herring (15.6%) and Sardine (14%) are the most common species, followed by Salmon (13.6%), Cod (12.4%), and Tuna and Trout (both at 11.6%). Other species like Shark (11%) and Snapper (10.4%) have slightly lower representation.

Finally, for the Water Pollution Level, 37% of the cases are from the category of Medium, followed by High (33.4%), and Low is 29.6%. This distribution implies that many regions have at least a medium level of pollution.

This gives a view of the overall dataset and indicates areas where more detailed analysis and interventions might be necessary, especially about overfishing risks and pollution levels.

The analysis focuses on the understanding of key variables related to marine fish populations, including Fish Population, Average Size, and Water Temperature. Descriptive statistics were used to explore the central tendencies, variability, and distributional properties of these variables. This gives a background for further investigation into fish population dynamics and environmental conditions.

Table 2 presents the results of descriptive statistics analysis for Fish Population, Average Size (cm), and Water Temperature (°C). From the sample size of 500, the mean of Fish Population is 5628.79 with a standard deviation of 2541.52, ranging from 1004 to 9996. The skewness value of -0.094 and kurtosis value of -1.18 indicate that the data is approximately symmetric and the distribution is relatively flat, suggesting no significant non-normality.

The Average Size of fish has a mean of 56.02 cm with a standard deviation of 24.96, ranging from 10.23 cm to 99.79 cm; the skewness is -0.07 and kurtosis -1.11, hence closely approaching normality with no reliable skew or extreme peaks. So does Water Temperature with its average at 22.55°C ±4.36°C ranging from 15-29.99°C. The skewness value of -0.078 and kurtosis value of -1.13 confirm that this variable also follows a distribution close to normal. These results show that the three variables are well-distributed, with no noticeable outliers or heavy non-normality issues.

The analysis has shown that the distribution of Fish Population, Average Size, and Water Temperature is appropriate-there is no substantial skewness or kurtosis to show non-normality. Such results prove the dependability of the dataset for further research and give a view on how stable the distribution of fish characteristics and their environmental factors is.

**Table-2 Descriptive statistics**



Before performing the correlation analysis, the normality of the variables Fish Population, Average Size (cm), and Water Temperature (°C) was investigated by the Shapiro-Wilk and Kolmogorov-Smirnov tests. It is observed from Table 3 that all three variables are significantly deviated from normality, as both tests have p-values less than 0.001, hence rejecting the null hypothesis of normality.

Concretely, the Shapiro-Wilk statistic for Fish Population, Average Size in cm, and Water Temperature in °C was 0.956, 0.962, and 0.956, respectively. Considering these, nonparametric tests, such as Spearman's correlation, will be recommended. Alternatively, data transformations may be considered to approximate normality in order to consider applying tests that require it.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table-3 Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | |  |
|  | Statistic | df | Sig. | Statistic | df | Sig. |
| Fish\_Population | 0.069 | 500 | <.001 | 0.956 | 500 | <.001 |
| Average\_Size(cm) | 0.073 | 500 | <.001 | 0.962 | 500 | <.001 |
| Water\_Temperature(C) | 0.063 | 500 | <.001 | 0.956 | 500 | <.001 |

**Spearman's Correlation Analysis**

We have performed a correlation analysis among Fish Population, Average Size (cm), and Water Temperature (°C). Before performing the correlation analysis, a normality test was conducted, and it showed that the data was not normally distributed. Hence, Spearman's Rank-Ordered Correlation (a non-parametric method) was used to check the relationship between the variables.

The results of the Spearman's correlation analysis are shown in Table 2. From this analysis, it can be seen there is a positive, weak, and statistically significant association between Fish Population and Water Temperature (r = 0.105, p = 0.019). This suggests that as water temperature rises, fish population slightly increases.

There is also a negligible and non-significant relationship between Fish Population and Average Size, with r = 0.087 and p = 0.052, indicating no effective relationship between these variables. In the same way, the relationship of Average Size and Water Temperature is negligible and nonsignificant, with the correlation coefficient r = -0.030 and p-value = 0.502, showing no notable relationship between these variables.

These results indicate that the rest of the relationships among the variables are negligible and statistically insignificant, except for a minor effect of water temperature on fish population. Future studies could investigate nonlinear or contextual influences on these relationships.

**Table-4 Spearman’s correlation analysis**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **Variables** | **1** | **2** | **3** | | 1) Fish Population | 1 | 0.087 | 0.105 | | 2) Average Size (cm) | 0.087 | 1 | -0.03 | | 3) Water Temperature (°C) | 0.105 | -0.03 | 1 | | Significant at p < 0.05. | |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

**Cross Tabulation Analysis**

Cross-tabulation of Fishing Method by Region was performed. Such an analysis will provide the distribution of fishing methods across regions. The results are shown in the table below, displaying counts and percentages by fishing method and region.

In this regard, according to the result, the most widely used gear types throughout the regions are Net fishing and Line fishing with 31% and 34.4%, respectively. On the other hand, Line fishing is highest in the Pacific Ocean at 41%, while Net fishing is highest in the North Atlantic at 36.7%. Not surprisingly, Trawl fishing also makes a high percentage of 30.2% of the total, with the highest concentration in the Pacific Ocean at 34.6%.

The Indian Ocean, for example, exhibits a singular behavior where traditional fishing methods represent 100% of the observations, without the usage of Line, Net, or Trawl fishing. It presents a marked regional feature of the fishing methods applied.

Cross-tabulation analysis indicates large variances of fishing methods across regions: line fishing in the Pacific Ocean and Net fishing in the North Atlantic. Interestingly, the Indian Ocean relies solely on traditional methods of fishing. From here, regional preferences and differences show up, which may be used to develop sustainable fishing policies.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Table-5 Fishing\_Method Region Crosstabulation** | | | | | | |
|  |  | Region |  |  |  |  | Total |
|  |  |  | Indian Ocean | Miditerranena Sea | Noth Atlantic | Pacific Ocean | |
| Fishing\_Method | Count | 23 | 0 | 0 | 0 | 0 | 23 |
|  | % within Fishing\_Method | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% |
|  | % within Region | 100.00% | 0.00% | 0.00% | 0.00% | 0.00% | 4.40% |
| Line | Count | 0 | 39 | 46 | 48 | 47 | 180 |
|  | % within Fishing\_Method | 0.00% | 21.70% | 25.60% | 26.70% | 26.10% | 100.00% |
|  | % within Region | 0.00% | 36.10% | 33.10% | 41.00% | 34.60% | 34.40% |
| Net | Count | 0 | 37 | 51 | 32 | 42 | 162 |
|  | % within Fishing\_Method | 0.00% | 22.80% | 31.50% | 19.80% | 25.90% | 100.00% |
|  | % within Region | 0.00% | 34.30% | 36.70% | 27.40% | 30.90% | 31.00% |
| Trawl | Count | 0 | 32 | 42 | 37 | 47 | 158 |
|  | % within Fishing\_Method | 0.00% | 20.30% | 26.60% | 23.40% | 29.70% | 100.00% |
|  | % within Region | 0.00% | 29.60% | 30.20% | 31.60% | 34.60% | 30.20% |
| Total | Count | 23 | 108 | 139 | 117 | 136 | 523 |
|  | % within Fishing\_Method | 4.40% | 20.70% | 26.60% | 22.40% | 26.00% | 100.00% |
|  | % within Region | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |

Analysis of Fish Population Across Regions

This analysis will explain whether the number of Fish Population significantly varies across the four regions: Indian Ocean, Mediterranean Sea, North Atlantic, and Pacific Ocean. Before the comparison is made, the data will be checked for normality using the Shapiro-Wilk test. If normality is assumed, a parametric test such as One-Way ANOVA will be applied. If normality is not met, then a non-parametric alternative will be performed using the Kruskal-Wallis test to determine differences across the regions.

Fish Population was compared across four regions: Indian Ocean, Mediterranean Sea, North Atlantic, and Pacific Ocean. The objective of this analysis was to determine if the Fish Population is significantly different among the regions. The Shapiro-Wilk test results indicated that Fish Population in all four regions was not normally distributed since the p-values for all groups were less than 0.05. Since the normality assumptions have been violated, the hypothesis is tested using the Kruskal-Wallis Test.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | | |  |
|  | Region | Kolmogorov-Smirnovb | | Shapiro-Wilk |  | |  |
|  |  | Statistic | Sig. | Statistic | df | Sig. |  |
| Fish\_Population | Indian O | 0.094 | 0.02 | 0.937 | 108 | <.001 |  |
|  | Mediterr | 0.083 | 0.019 | 0.96 | 139 | <.001 |  |
|  | North At | 0.088 | 0.026 | 0.963 | 117 | 0.003 |  |
|  | Pacific | 0.088 | 0.011 | 0.947 | 136 | <.001 |  |

The results of the Kruskal-Wallis test indicated no statistically significant difference in Fish Population across the four regions, χ²(3) = 2.752, p = 0.432. The mean rank of Fish Population for the Indian Ocean was 250.12, the Mediterranean Sea was 260.34, the North Atlantic was 254.89, and the Pacific Ocean was 256.45. Kruskal-Wallis test showed that Fish Population did not differ among the four regions statistically: χ²(3) = 2.752, p = 0.432. This suggests that Fish Population has no significant variation between regions; this could mean no major variation in Fish Population between regions across the dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table-7** | **Independent-Samples Kruskal-Wallis Test** | |  |
| Total N | 500 |  |  |
| Test Statistic | 2.752a |  |  |
| Degree Of Freedom | 3 |  |  |
| Asymptotic Sig.(2-sided test) | 0.432 |  |  |

The following research examines the relationship between Average Size (cm), Water Temperature (C), and Fish Population through multiple regression analysis. This will help in identifying the significance of independent variables in predicting the variation in fish population. Before conducting the analysis, assumptions like multicollinearity and model fit were checked to ensure that the results obtained are valid.

A regression was carried out using Fish Population as the dependent variable and Average Size-cm and Water Temperature-C as independent variables. The model presented an R value of 0.135, indicating small positive relationship between predictors and the dependent variable. In turn, it has an R² value of 0.018, indicating that the explained variance in fish population by predictors is 1.8%. The adjusted R², accounting for possible overfitting, was 0.014. Although the percentage variance explained by the model was small, it still turned out to be significant due to an F-statistic of 4.635 and a p-value of 0.010 less than 0.05 obtained from ANOVA. This confirms that the independent variables collectively have a significant relationship with Fish Population.

The coefficients are of further use in highlighting the contributions of each independent variable. The constant term, with an intercept of 3773.678, represents the expected baseline fish population when all predictors are zero. This value was highly significant (p < 0.001). Average Size (cm) was positively related to fish population; for every one-unit increase in average size, there was an increase in the population of 9.450, assuming other conditions remain constant. This is an outstanding relationship since the p-value has a significance probability of 0.037. Moreover, Water Temperature (C) was also positively related to fish population; for each additional unit of temperature, the population rises by 58.784 units. This variable was also significant at p = 0.024.

Collinearity diagnostics did not raise any multicollinearity issues, with the VIF values for the two predictors estimated at 1.001 each, far below the commonly accepted cut-off point of 10. The following is the regression equation obtained based on the analysis:

**Fish Population (Predicted) = 3773.678 + 9.450(Average Size (cm) + 58.784(Water Temperature (C)**

From these results, it becomes evident that Average Size, cm, and Water Temperature, C, are both strong causes of fish population. From these two predictors, water temperature, C, resulted in a higher effect than fish population, with a higher coefficient value of 58.784, while the coefficient for average size was 9.450. However, the overall model explained only a small portion of the variance in fish population as indicated by the relatively low R² value. This would then imply that other factors outside of those fitted in the model might contribute to the determination of fish population.

Conclusion

In conclusion, from the results of multiple regression analysis, it appears that Average Size (cm) and Water Temperature (C) are the significant predictors of fish population. Although the model has a low explanatory power, it gives useful insights into the relationship between these ecological variables. Further research can improve the model by adding more predictors or considering interaction effects to understand the factors affecting fish population. This analysis forms the basis for ecological management and conservation strategies that will help in sustaining fish populations.

|  |  |  |
| --- | --- | --- |
| **Independent Variables** | **Model** | **Sig** |
| **Average Size (cm)** | 9.45(2.088) | *0.037* |
| **Water Temperature (°C)** | 58.784(2.268) | *0.024* |
| **Constant** | 3773.678(5.781) | *<0.01* |
| **F-statistic** | 4.635 |  |
| **Adjusted R²** | 0.014 |  |
| **N (Sample Size)** | 500 |  |

Logistic regression analysis was conducted to predict the likelihood of Overfishing Risk, which is coded as 1 for "Yes" and 0 for "No," from the predictors: Fishing Method, Fish Population, Average Size in cm, Water Temperature in C, and Water Pollution Level. This analysis evaluated the significance and effect of the independent variables on the dependent variable.

Results and Analysis

Logistic regression analysis was used to assess the impact of Fishing Method, Fish Population, Average Size (cm), Water Temperature (C), and Water Pollution Level on Overfishing Risk. The performance of the model is summarized in Table 1. The -2 Log Likelihood value of 679.525 and Nagelkerke R² of 0.035 indicated that only 3.5% of the variance in Overfishing Risk was explained by the model. The result from the Likelihood Ratio Test was χ² = 13.422, df = 7, p = 0.062, which is not significant. Hence, this would suggest that, overall, the model was not a good fit for predicting Overfishing Risk at 0.05 significance level.

Examining the predictors individually, none were statistically significant. Fishing Method (1), OR = 1.392, Wald χ² = 2.259, p = 0.133 and Fishing Method (2), OR = 0.949, Wald χ² = 0.054, p = 0.816 did not reach statistical significance to predict Overfishing Risk. Therefore, neither fishing method was a strong predictor of overfishing.

Another nonsignificant relation was that of the Fish Population predictor: OR = 1, Wald χ² = 2.227, p = 0.136. Although the odds ratio was indicates no significant raise of risk, the lack of significance implies that changes in the fish population variability alone do not explain changes in the probability of overfishing. Similarly, Average Size (cm) (OR = 1.006, Wald χ² = 2.740, p = 0.098) and Water Temperature (C) (OR = 1.036, Wald χ² = 2.772, p = 0.096) did not exhibit a statistically significant relationship with Overfishing Risk, even though their odds ratios were slightly above 1, indicating a trend of increasing risk with higher size and temperature, albeit non-significant.

The categorical variable Water Pollution Level was explored at two levels. Both Water Pollution Level (1) with OR = 1.23, Wald χ² = 0.812, p = 0.368, and Water Pollution Level (2), with OR = 0.947, Wald χ² = 0.063, p = 0.802, did not show significant relations with Overfishing Risk. This probably means that the present categorization of water pollution does not strongly affect the risk of overfishing.

Interestingly, the constant term was statistically significant-OR = 0.2, Wald χ² = 8.202, p = 0.004-indicating a baseline level of risk to overfishing in the absence of the predictors. This underlines the fact that probably external factors or factors not measured may influence overfishing risk.

Overall, the predictors involved in this model were not the significant contributors towards the prediction of Overfishing Risk. Though some trends were identified, such as a potential increase in risk with higher temperatures and larger fish sizes, these did not reach statistical significance. This is further driven home by the low variance explained by the model, Nagelkerke R² = 0.035, implying that other critical variables that drive Overfishing Risk were not embraced in this analysis.

In the logistic regression, none of the Fishing Method, Fish Population, Average Size (cm), Water Temperature (C), and Water Pollution Level predictors proved to be significant to Overfishing Risk. Only the constant showed the base level of risk, and the model was weakly explanatory in general. These results would suggest further investigation with additional variables or alternative model structures for better analysis of factors influencing Overfishing Risk.

Table-8 Logistic regression prediction

|  |  |  |
| --- | --- | --- |
| **Independent Variables** | **Odds Ratio** | **p-value** |
| Fishing Method (1) | 1.392(2.259) | 0.133 |
| Fishing Method (2) | 0.949(0.054) | 0.816 |
| Fish Population | 1(2.227) | 0.136 |
| Average Size (cm) | 1.006(2.74) | 0.098 |
| Water Temperature (C) | 1.036(2.772) | 0.096 |
| Water Pollution Level (1) | 1.23(0.812) | 0.368 |
| Water Pollution Level (2) | 0.947(0.063) | 0.802 |
| **Constant** | 0.2(8.202) | 0.004 |
| −2LogLikelihood | 679.525 |  |
| Likelihood Ratio Test (χ2/df) | 13.422 / 7 |  |
| Nagelkerke R2 | 0.035 |  |
| *Wald-χ2\chi^2χ2 statistics are in parentheses.* | | |
|  |  |  |