ADSC1910 Case 6

Neha Malhan(T00757554) & Thiwanka Abeyweera(T00754871)

2024-10-22

Introduction

This study analyzes whether Northern Italian forest environments offer the best conditions for wild birds, as many ecologists believe. We seek to determine whether forests consistently support higher bird populations than other habitat types, such as agricultural or urban areas, by examining data on Bird Species Richness (BSR) and DOMinant Land Use (DOM.LU). By determining the best habitats for maintaining bird populations, this study will help guide conservation efforts by offering insightful information on the preferences of wild birds. We also discuss possible drawbacks, like the risks of performing many comparisons to make sure the conclusions are reliable.

Methods

In this study, Bird Species Richness (BSR) data were first analysed using descriptive statistics, including mean, median, and standard deviation, along with visualizations such as box plots and histograms to compare forest and non-forest habitats. These steps provided a clear view of the distribution of bird populations across the different habitat types. The following are the hypothesis formulations for this analysis:

Null Hypothesis $H_0 : \mu_{forest} - \mu_{non\text{-}forest} \leq 0$,indicating that forest habitats support the same or fewer bird species compared to non-forest habitats.

Alternative Hypothesis $H_1 : \mu_{forest} - \mu_{non\text{-}forest} > 0$, suggesting that forest habitats support significantly more bird species.

A one-tailed two-sample t-test was performed to determine whether forest habitats support significantly higher bird populations. The descriptive analysis complemented the hypothesis test by providing context for the statistical findings.

Findings/Result

1) Descriptive Summary of Bird Species Richness in Forest and Non-Forest Habitats

Descriptive statistics were generated to summarize bird species richness in forest and non-forest habitats, highlighting key trends in the data.

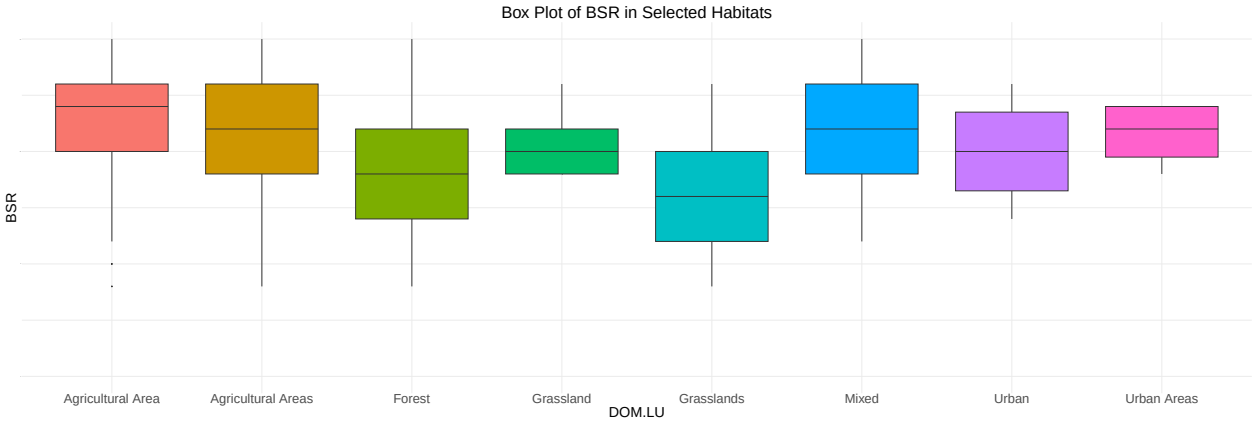
Table 1: Descriptive Statistics for Forest and Non-Forest Habitats:

Habitat	Mean_BSR	Median_BSR	SD_BSR
Forest	22.99464	11	31.81482
Non-Forest	18.01501	12	36.76618

The analysis shows that forest habitats generally contribute a higher average bird species richness than non-forest habitats, although the median richness is slightly lower for forests. The variation in bird species richness, as indicated by the standard deviation, is significant for both habitat types, with non-forest habitats showing a wider spread. This suggests that while forests may support more bird species on average, the distribution of species richness varies considerably across both habitat types.

2) Visualization of Bird Species Richness Across Habitat Types

The plot provides insights into how bird species richness varies across different environments, highlighting the differences in species distribution and abundance among the habitats analyzed. For better visualization and clarity, two categories with lesser significance were removed, as they had only a single observation each.



The plot compares bird species richness across different habitat types, including forests, agricultural areas, grasslands, and urban areas. Forest habitats show a wider range of bird populations, with more sites having higher species counts compared to other habitat types. While some habitats, like urban areas, tend to have lower and more consistent bird species richness, forests exhibit a more varied distribution with some locations supporting significantly more birds.

3) Statistical Analysis of Bird Species Richness Between Forest and Non-Forest Habitats

The two Sample t-test was conducted to compare the average bird species richness between forest and non-forest habitats. The test statistic ($t = 2.1733$) indicates how far the observed difference in means is from zero, with a larger value showing a greater difference between the two groups. The degrees of freedom ($df = 864.35$) represent the amount of information used to estimate this difference, adjusting for sample sizes and variances. The p-value ($p = 0.01501$) is less than the common significance level of 0.05, providing enough evidence to reject the null hypothesis. The mean of x (22.99) represents the average bird species richness in forest habitats, while the mean of y (18.02) represents the average richness in non-forest habitats. These values suggest that forest habitats generally support a higher number of bird species compared to non-forest habitats, and the t-test confirms that this observed difference is statistically significant.

Conclusion

The analysis shows that forest habitats generally support a higher average number of bird species compared to non-forest habitats, despite having a slightly lower median and more variation in bird counts. The visualizations demonstrate that forests have a wider range of bird populations, with more sites having higher species counts, while non-forest habitats show more consistent but lower species richness. The two Sample t-test indicates that the difference in bird species richness between the two habitats is statistically significant, leading us to reject the null hypothesis in favor of the alternative hypothesis. This confirms that forest habitats are more favorable for bird species richness and highlights the importance of conserving these areas to support bird populations.

Recommendation

Based on these findings, it is recommended to focus on protecting and restoring forest habitats to help support more bird species. Conservation efforts should prioritize preserving forests and planting new trees. More research could also help understand which features of forest habitats make them more favorable for bird species, leading to better conservation actions.

References

[1] Hellingman, S. (2024). Hypothesis Testing 1 and Hypothesis Testing 2. ADSC 1000, Fall 2024.

Appendices

This case study was collaboratively developed by Neha Malhan and Thiwanka Abeyweera.

```
knitr::opts_chunk$set(echo = TRUE)

# Load necessary libraries
library(ggplot2)
library(knitr)
library(dplyr)


# Reading the dataset
IB_data <- read.csv("Italian_Birds.csv")

# Dividing data into two groups
forest_BSR <- data.frame(BSR = IB_data[IB_data$DOM.LU == 'Forest', 'BSR'])
non_forest_BSR <- data.frame(BSR = IB_data[IB_data$DOM.LU != 'Forest', 'BSR'])

# Calculating Descriptive Statistics
forest_stats <- forest_BSR %>%
  summarise(
    Habitat = "Forest",
    Mean_BSR = mean(BSR, na.rm = TRUE),
    Median_BSR = median(BSR, na.rm = TRUE),
    SD_BSR = sd(BSR, na.rm = TRUE)
  )
```

```r
non_forest_stats <- non_forest_BSR %>%
  summarise(
    Habitat = "Non-Forest",
    Mean_BSR = mean(BSR, na.rm = TRUE),
    Median_BSR = median(BSR, na.rm = TRUE),
    SD_BSR = sd(BSR, na.rm = TRUE)
  )

# Displaying the summary
kable(bind_rows(forest_stats,non_forest_stats), caption = "Descriptive Statistics for
                                                           Forest and Non-Forest Habitats:")
```

```r
# Excluding two categories having lesser observations
Boxplot_data <- IB_data %>%
  filter(!DOM.LU %in% c("River", "Riparian Vegetation or Reforestation"))

# Create the boxplot with filtered data
ggplot(Boxplot_data, aes(x = DOM.LU, y = BSR, fill = DOM.LU )) +
  geom_boxplot() +
  labs(title = "Box Plot of BSR in Selected Habitats",
       x = "DOM.LU",
       y = "BSR") +
  ylim(0, 15) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 90),       # Larger x-axis labels
    axis.title.x = element_text(size = 100),      # Larger x-axis title
    axis.title.y = element_text(size = 100),      # Larger y-axis title
    plot.title = element_text(size = 115, hjust = 0.5), # Larger plot title and centered
    legend.position = "none"
  )

# Performing one-tailed two-sample t-test
t_test <- t.test(forest_BSR, non_forest_BSR, alternative = "greater")

# Printing the t-test results
print(t_test)
```

Below is the two-sample t test result:

```r
# Performing one-tailed two-sample t-test
t_test <- t.test(forest_BSR, non_forest_BSR, alternative = "greater")

# Printing the t-test results
print(t_test)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  forest_BSR and non_forest_BSR
## t = 2.1733, df = 864.35, p-value = 0.01501
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.206843      Inf
## sample estimates:
## mean of x mean of y
##  22.99464  18.01501
```