

Descriptive and Univariate Vibe Analysis of Forestry Data with AI and R Statistics

Kato Samuel Namuene¹; Ndinge Nadia Mbella²

¹Department of Forestry and Wildlife, Faculty of Agriculture and Veterinary Medicine, University of Buea, Buea, Cameroon

²Department of English Law, Faculty of Laws and Political Science, University of Buea, Buea, Cameroon

Publication Date: 2026/04/14

Abstract: This study presents a reproducible, AI-assisted framework for descriptive and univariate statistical analysis of ecological count data, integrating vibe data analysis with conventional manual methods using snapping frequency observations from 94 tree species in Korup National Park, Cameroon. Using Claude.ai to generate R statistical code through structured prompt engineering, we systematically compare classical parametric approaches (t-test, Z-test) with non-parametric alternatives (Wilcoxon signed-rank test, sign test, bootstrap confidence intervals) to determine the most appropriate analytical framework for forestry count data across four stages; exploratory data analysis, normality assessment, hypothesis testing, and outlier detection. Snapping frequency exhibited extreme positive skewness (5.087) and leptokurtic distribution (kurtosis = 36.725), with all four normality tests, Shapiro-Wilk ($W = 0.524$), Anderson-Darling ($A = 10.109$), Kolmogorov-Smirnov ($D = 0.267$), and D'Agostino-Pearson ($K^2 = 135.50$) unanimously rejecting normality (all $p < 0.0001$). The mean snapping frequency was 1.333 (SD = 1.958), with median and mode equal to 1.0, and 84.9% of species recording two or fewer events. Non-parametric tests confirmed the population median did not significantly differ from 1.0 (Wilcoxon $p = 0.180$; sign test $p = 1.000$; bootstrap 95% CI: [1.0, 1.0]), while all tests confirmed the median exceeded zero ($p < 0.0001$). Two outlier species were unanimously identified across all three detection methods: *Oubanguia alata* (16 snappings; $z = 7.49$) and *Protomegabaria stapfiana* (8 snappings; $z = 3.41$). Comparison of vibe analysis and manual analysis across 16 statistical outputs revealed complete numerical equivalence, with the AI demonstrating autonomous assumption-aware method selection without explicit instruction. While vibe analysis completed all stages within a single iterative session, mandatory validation through executed R code and analyst oversight remain essential. This framework provides forestry researchers with accessible, validated tools for rigorous, reproducible statistical analysis of non-normal count data.

Keywords: Descriptive Statistics; Univariate Analysis; Parametric Tests; Non-Parametric Tests; Normality Testing; Artificial Intelligence; R Programming; Data Visualization; Ecological Data Analysis.

How to Cite: Kato Samuel Namuene; Ndinge Nadia Mbella (2026) Descriptive and Univariate Vibe Analysis of Forestry Data with AI and R Statistics. *International Journal of Innovative Science and Research Technology*, 11(4), 513-540. <https://doi.org/10.38124/ijisrt/26apr077>

I. INTRODUCTION

The selection of appropriate statistical methods is fundamental to forestry research and data science, directly influencing the validity, reliability, and reproducibility of scientific conclusions (Chaoubi and Djalab, 2025; Miller and Spiegel, 2025). Descriptive statistics form the foundational step in any quantitative forestry investigation, providing essential summaries of central tendency, dispersion, and distributional shape that guide subsequent analytical decisions (Zar, 2010). These preliminary analyses are indispensable for characterizing stand structure, species diversity metrics, and dendrometric variables before inferential testing (Husch et al., 2003). Univariate analysis extends this exploration by examining individual variables in isolation, enabling

researchers to detect patterns, anomalies, and departures from normality within forest mensuration datasets (Sokal & Rohlf, 2012). Collectively, descriptive and univariate procedures establish the empirical foundation upon which robust parametric or non-parametric test selection is justified (Field, 2018).

Parametric tests, which rely on specific distributional assumptions, offer greater statistical power and efficiency when their underlying assumptions are satisfied (Shukla et al., 2025). However, forestry data frequently violate these assumptions, necessitating the use of non-parametric alternatives that provide robustness in the face of non-normality, heteroscedasticity, and outliers (Potvin & Roff, 1993; Underwood, 1997). This methodological dichotomy

presents researchers with a critical decision-making challenge that has profound implications for statistical inference and the interpretation of ecological phenomena (Warton & Hui, 2011).

In forestry studies, particularly those involving count data such as species behavior frequencies, foraging patterns, or reproductive outputs, data often exhibit marked departures from normality characterized by skewness, excess kurtosis, zero-inflation, and the presence of extreme values (O'Hara & Kotze, 2010; Warton et al., 2017). Traditional parametric approaches, including Student's t-test and analysis of variance (ANOVA), assume that data follow a normal distribution with homogeneous variance across groups assumptions that are frequently violated in natural systems (Crawley, 2013; Logan, 2012). When these assumptions are violated, parametric tests may produce inflated Type I error rates, reduced statistical power, or fundamentally misleading conclusions about ecological relationships (Underwood, 1997; McDonald, 2014). Conversely, non-parametric methods, often termed "distribution-free" tests, make minimal assumptions about the underlying data distribution and instead rely on rank transformations or permutation-based inference (Hollander et al., 2015; Bonnini et al., 2024). These methods can provide more reliable and valid inferences when data deviate substantially from normality, though they may sacrifice some statistical power when parametric assumptions are met (Conover & Iman, 1981; Fagerland & Sandvik, 2009; Bonnini et al., 2024).

The advancement of computational tools and statistical software has democratized access to sophisticated analytical methods, yet the proliferation of analytical options has also increased the complexity of methodological decision-making (Hampton et al., 2015; Poisot et al., 2019). R statistical programming, an open-source language specifically designed for statistical computing and graphics, has emerged as the dominant platform for ecological data analysis due to its flexibility, extensive package ecosystem, and strong community support (Wickham et al., 2019; R Core Team, 2025). However, the effective use of R requires not only statistical knowledge but also programming proficiency, creating barriers for researchers without formal computational training (Baumer et al., 2014; Touchon & McCoy, 2016). This paper demonstrates a systematic approach to descriptive and univariate data analysis using both parametric and non-parametric methods, implemented through reproducible R code. We analyze snapping behavior data from 94 tree species to illustrate the complete analytical workflow, from initial data exploration and visualization through formal hypothesis testing and comparative method evaluation. Our objective is to provide researchers with accessible code templates, interpretative guidance, and a structured decision framework for selecting appropriate analytical approaches based on empirical data characteristics.

The integration of artificial intelligence (AI) in statistical programming represents a paradigm shift in data analysis, offering unprecedented opportunities to enhance research efficiency, reduce coding errors, and improve methodological rigor (Beam & Kohane, 2018; Topol, 2019). Large language

models (LLMs) and AI-assisted programming tools have demonstrated remarkable capabilities in generating statistical code, debugging analyses, and providing methodological recommendations tailored to specific research contexts (Chen et al., 2021; Vaithilingam et al., 2022). This technological evolution has the potential to bridge the gap between statistical theory and practical implementation, making advanced analytical techniques accessible to researchers across disciplines and career stages (Lowndes et al., 2017). By leveraging AI-generated R code alongside traditional statistical expertise, researchers can accelerate their analytical workflows while maintaining scientific rigor and reproducibility. This convergence of AI technology and statistical programming offers a promising pathway to address long-standing challenges in ecological data analysis, including the selection of appropriate methods, validation of statistical assumptions, and interpretation of complex results (Recknagel, 2001; Thessen, 2016).

➤ *Artificial Intelligence in Statistical Analysis*

Artificial intelligence has fundamentally transformed the landscape of scientific computing and statistical analysis over the past decade (Esteva et al., 2019; Senior et al., 2020). The development of large language models, including GPT-3, GPT-4, and Claude, has created powerful tools capable of understanding natural language queries and generating sophisticated programming code across multiple languages (Brown et al., 2020; OpenAI, 2023). In the context of statistical analysis, AI systems can interpret research questions formulated in plain language, recommend appropriate analytical methods based on data characteristics, generate executable code, and provide interpretative guidance for statistical output (Sarker, 2022; Kasneci et al., 2023). These capabilities address several critical challenges in contemporary ecological research, including the technical barrier to entry for complex statistical methods, the time-intensive nature of code development and debugging, and the difficulty of staying current with rapidly evolving methodological best practices (Hampton et al., 2017; Gomes et al., 2022).

The application of AI in statistical programming extends beyond simple code generation to encompass more sophisticated tasks such as automated data cleaning, exploratory data analysis, assumption checking, and results interpretation (Stoyanovich et al., 2020; Wang et al., 2021). AI systems can analyze data structure, identify potential violations of statistical assumptions, and suggest appropriate transformations or alternative analytical approaches (Xu et al., 2021). Furthermore, AI-assisted programming can enhance reproducibility by generating well-documented, standardized code that follows current best practices in statistical programming (Peng, 2015; Sandve et al., 2013). This capability is particularly valuable in ecological research, where complex datasets with multiple covariates, hierarchical structures, and non-standard distributions are commonplace (Bolker et al., 2009; Harrison et al., 2018). However, the integration of AI in statistical analysis also raises important questions about validation, interpretability, and the appropriate balance between automation and human expertise (Rudin, 2019; Selbst et al., 2019).

- *AI Prompts for Statistical Programming*

The effectiveness of AI-assisted statistical programming depends critically on the quality and specificity of user prompts the natural language instructions provided to the AI system (White et al., 2023; Zhou et al., 2023). Well-constructed prompts should clearly articulate the research objective, describe the data structure and characteristics, specify the desired analytical approach, and indicate any constraints or preferences regarding output format (Liu et al., 2023; Zamfirescu-Pereira et al., 2023). For statistical analysis in R, effective prompts typically include information about the response and predictor variables, sample size, data distribution, the research hypothesis, and specific visualization or output requirements (Sarker, 2022). For example, a comprehensive prompt might specify: "Generate R code to compare snapping behavior frequencies between two tree species groups using both parametric (t-test) and non-parametric (Mann-Whitney U test) approaches. Include normality testing, descriptive statistics, appropriate visualizations, and interpretation guidelines." Such detailed prompts enable AI systems to generate more accurate, context-appropriate code that addresses the specific analytical needs of the research question (Reynolds & McDonnell, 2021; Vaithilingam et al., 2022).

The iterative refinement of AI prompts represents a crucial skill for researchers seeking to leverage these tools effectively (Singla et al., 2023). Initial prompts may produce code that requires modification to address specific data structures, analytical nuances, or output preferences (Barke et al., 2023). Through successive prompt iterations, researchers can guide the AI system toward increasingly refined solutions that precisely match their analytical requirements (White et al., 2023). This iterative process also serves an educational function, helping researchers develop deeper understanding of statistical methods and programming syntax through active engagement with AI-generated code (MacNeil et al., 2023; Denny et al., 2024). However, researchers must maintain critical oversight of AI-generated code, validating results against statistical theory, checking for logical errors, and ensuring that the analytical approach aligns with the specific characteristics of their data and research questions (Liao & Vaughan, 2023). The most effective use of AI in statistical programming combines the efficiency and accessibility of automated code generation with the domain expertise, critical thinking, and statistical knowledge of trained researchers (Sarker, 2022; Zhou et al., 2023).

- *R Statistical Programming for Forestry Analysis*

R has established itself as the preeminent statistical programming environment for ecological research, offering a comprehensive ecosystem of packages and tools specifically designed for environmental and biological data analysis (Crawley, 2013; Davis and Kay, 2023). Initially released in 1995 as an open-source implementation of the S programming language, R has evolved into a sophisticated platform supporting virtually every statistical method used in contemporary ecological research, from basic descriptive statistics to complex hierarchical models, spatial analysis, and machine learning approaches (Ihaka & Gentleman, 1996; R Core Team, 2025). The language's flexibility allows

researchers to implement custom analytical workflows, develop reproducible analysis pipelines, and create publication-quality visualizations through packages such as ggplot2, which implements the grammar of graphics framework (Wickham et al., 2015; Cox, 2007). The open-source nature of R has fostered a vibrant global community of developers and users who continuously contribute new packages, methodological innovations, and educational resources (Tippmann, 2015; Wickham et al., 2019).

The R ecosystem includes numerous specialized packages for ecological data analysis, each addressing specific analytical challenges common in environmental research (Borcard et al., 2018; Oksanen et al., 2025). For univariate analysis and hypothesis testing, packages such as stats (core R functionality), car (Companion to Applied Regression), and agricolae provide comprehensive implementations of both parametric and non-parametric methods (Fox & Weisberg, 2019; de Mendiburu, 2023). Data manipulation and cleaning are facilitated through the tidyverse suite of packages, including dplyr for data transformation and tidyr for data reshaping (Wickham et al., 2019). Assumption checking and diagnostic testing are supported through packages such as nortest for normality testing and car for homogeneity of variance assessment (Gross & Ligges, 2015; Fox & Weisberg, 2019). The integration of these packages within a unified programming environment enables researchers to develop complete analytical workflows that progress seamlessly from data import and cleaning through statistical analysis, visualization, and results reporting (Wickham, 2014; Gandrud, 2018).

Despite its powerful capabilities, R presents a significant learning curve for researchers without prior programming experience, potentially limiting its accessibility to those who would benefit most from its analytical power (Baumer et al., 2014). Common challenges include mastering R's syntax and data structures, understanding package installation and management, debugging code errors, and selecting appropriate functions from the vast array of available options (Touchon & McCoy, 2016). These barriers are particularly pronounced for students and early-career researchers who may lack formal training in computational methods (Lai et al., 2019). However, the combination of improved educational resources, user-friendly interfaces such as RStudio, and the emergence of AI-assisted programming tools has begun to address these accessibility challenges (Baumer et al., 2014; Vaithilingam et al., 2022). This democratization of R programming through AI assistance represents a significant opportunity to expand the research community's capacity for rigorous statistical analysis while maintaining the reproducibility and transparency that are hallmarks of R-based workflows (Hampton et al., 2017; Powers & Hampton, 2019).

- *Vibe Data Analysis: Modern Approaches to Statistical Practice*

Statistical practice is experiencing a fundamental shift with the emergence of AI-assisted approaches to data analysis. Large language models now enable what has been termed "vibe data analysis" (Mollick & Mollick, 2023) an approach emphasizing conceptual reasoning and iterative

exploration rather than rigid procedural workflows. Rooted in Tukey's (1977) exploratory data analysis and extending Breiman's (2001) critique of traditional "data modeling culture," this methodology allows researchers to express analytical intentions in natural language while receiving executable code and interpretive guidance (Breiman, 2001; Tukey, 2008). This shift supports recent calls for more transparent and accessible research practices (Nosek et al., 2015) by reducing syntactic burden and allowing analysts to focus on substantive questions such as patterns, relationships, and assumptions (Rane, 2023). However, meaningful use still requires statistical literacy to evaluate outputs and avoid misinterpretation (Gelman & Loken, 2019).

The approach is characterized by iterative refinement, rapid prototyping, and enhanced documentation through carefully constructed prompts. Researchers engage in conversational cycles that mirror expert analytical reasoning while making such processes accessible to a broader audience (Wickham & Golemund, 2016; Antoch et al., 2017). The quality of results depends critically on prompt design clear, well-structured queries yield more accurate and contextually appropriate statistical implementations (Denny, 2024). Multiple analytical strategies can be tested efficiently through prompt refinement, particularly during exploratory phases (Blei & Smyth, 2017), and conversational records capture methodological decisions that traditionally remain implicit, supporting transparency and reproducibility when appropriately curated (Jupyter et al., 2018). In practice, analysts pose questions in natural language, receive R or Python implementations, execute models, inspect results, and refine their approach through iterative prompting (Sarkar et al., 2022). This workflow is especially valuable in complex domains such as forestry research, where mixed models, machine learning techniques, and missing-data imputation must be integrated (van Buuren & Groothuis-Oudshoorn, 2011).

Despite these advantages, *vibe* data analysis introduces epistemological risks. The ease of iterative analysis may encourage *p*-hacking or hypothesizing after results are known (Head et al., 2015; Gelman & Loken, 2019), while the opaque reasoning of LLMs can obscure assumptions and methodological choices (Bommasani et al., 2021). Although automation may democratize research and accelerate discovery (Kitchin, 2014), critics warn it could weaken engagement with foundational statistical concepts (O'Neil, 2016). To mitigate these risks, researchers are advised to treat AI-generated outputs as provisional, critically review all code (Denny, 2024), preserve curated conversational records (Jupyter et al., 2018), maintain grounding in statistical principles (Sarkar et al., 2022), and use preregistration to distinguish confirmatory from exploratory analyses (Nosek et al., 2018). Looking forward, advances in LLMs may strengthen capabilities in causal inference and complex design handling (Pearl & Mackenzie, 2018), while integration into statistical education could enhance conceptual understanding by lowering technical barriers (Çetinkaya-Rundel & Rundel, 2018). Ultimately, *vibe* data analysis represents an evolution rather than a replacement of traditional statistical practice,

with its success dependent on balancing accessibility with rigor, transparency, and critical judgment.

➤ *Research Gaps and Objectives*

Despite widespread AI integration in statistical programming and extensive R adoption in forestry research, a critical gap exists in systematic guidance for assumption-aware univariate analysis of forestry behavioral data. Current resources either address parametric or non-parametric methods in isolation, or generate AI code without validation protocols for the distributional challenges common in tree species data; zero-inflation, extreme skewness, and outliers. This gap leaves researchers uncertain about method selection when normality assumptions fail, particularly when using AI tools that may produce syntactically correct but statistically inappropriate analyses.

Addressing this gap, the present study develops an AI-assisted analytical framework for forestry count data that:

- Demonstrates effective prompt engineering for generating assumption-aware R code tailored to ecological count data.
- Provides systematic protocols for normality assessment, outlier detection, and method selection.
- Delivers reproducible code templates for tree species behavioral observations.
- Explicitly compares parametric and non-parametric approaches using snapping behavior data from 94 species.
- Validates AI-generated outputs through multi-method assumption checking and visual diagnostics.
- Compares *vibe* analysis and manual analysis for numerical equivalence and methodological alignment.
- Establishes transparent decision criteria for selecting appropriate statistical methods based on empirical data characteristics.

➤ *Significance and Novelty*

This research advances AI-assisted forestry analytics by integrating large language model capabilities with rigorous statistical validation for ecological count data. The framework addresses three critical challenges: R programming complexity, non-normal data distributions in forestry research, and lack of transparent method selection criteria. The study's contributions include:

- Empirical validation: Analysis of real snapping frequency data from 94 tree species demonstrates the practical consequences of assumption violations on inferential outcomes.
- Methodological transparency: Side-by-side comparison of *t*-tests versus Wilcoxon tests reveals how normality violations affect inference ($p = 0.104$ vs. 0.180).
- Reproducible workflows: Complete R code organized as AI prompts spanning data exploration, assumption testing, hypothesis testing, and visualization.
- Decision frameworks: Explicit criteria for method selection based on normality tests, skewness thresholds, and data type.

- Educational value: Iterative prompt refinement demonstrates how to validate AI-generated code while building statistical literacy.

II. MATERIALS AND METHODS

This section details the data, software, and analytical framework used to investigate the distributional properties of snapping count observations recorded across 94 tree species in Korup National Park. The analysis was structured into four sequential stages; exploratory data analysis, normality testing, hypothesis testing, and outlier detection was conducted using R Statistical Software with AI-assisted code generation via Claude.ai (Anthropic, 2026).

A. Data Description

The dataset consists of snapping count observations for 94 tree species from Korup National Park. Snapping refers to the audible sound produced by certain tree species under specific environmental conditions. The data structure comprises two variables: Species (categorical identifier) and snapping (discrete count variable ranging from 0 to 16). This dataset represents a typical ecological count data scenario with potential overdispersion and non-normal distribution characteristics.

B. Statistical Software And Libraries

All analyses were performed using R Statistical Software. The following R packages were utilized:

- Base R: Core statistical functions including descriptive statistics, one-sample t-test, Kolmogorov-Smirnov test, Wilcoxon signed-rank test, Mann-Whitney U test, sign test (manual implementation), bootstrap resampling, and confidence interval estimation
- Ggplot2: Advanced data visualization
- Moments: Skewness, kurtosis calculations, and D'Agostino-Pearson normality test (via `agostino.test()`)
- Nortest: Additional normality testing, specifically the Anderson-Darling test (via `ad.test()`)
- BSDA: Basic Statistics and Data Analysis; used for the formal sign test (via `SIGN.test()`), with a base R manual fallback implemented where the package was unavailable

C. Analytical Framework

The analysis followed a systematic four-stage framework:

➤ Stage 1: Exploratory Data Analysis (EDA)

- Data structure examination and summary statistics
- Calculation of measures of central tendency (mean, median, mode)
- Calculation of measures of dispersion (variance, standard deviation, IQR, MAD)
- Calculation of measures of shape (skewness, kurtosis)

• Frequency Distribution Analysis

- ✓ Histogram with superimposed normal curve
- ✓ Bar chart of frequency distribution
- ✓ Box plot with individual data points

➤ Stage 2: Normality Testing

Multiple normality tests were applied to assess distributional assumptions:

- Shapiro-Wilk test: Most powerful for small to moderate sample sizes
- Anderson-Darling test: Gives more weight to the tails of the distribution
- Kolmogorov-Smirnov test: General goodness-of-fit test
- D'Agostino-Pearson test: Based on skewness and kurtosis
- Quantile-quantile (Q-Q) plot for normality assessment
- Kernel density plot

➤ Stage 3: Hypothesis Testing

Both parametric and non-parametric approaches were applied:

• Parametric Methods:

- ✓ One-sample t-test (two hypotheses: $\mu = 0$ and $\mu = 1$)
- ✓ Z-test (large sample approximation)
- ✓ Confidence intervals for the population mean

• Non-Parametric Methods:

- ✓ Wilcoxon signed-rank test
- ✓ Sign test
- ✓ Mann-Whitney U test
- ✓ Bootstrap confidence intervals for the median

➤ Stage 4: Outlier Detection

• Multiple Outlier Detection Methods were Employed:

- ✓ IQR method ($1.5 \times \text{IQR}$ rule)
- ✓ Z-score method ($|z| > 3$)
- ✓ Modified Z-score using median absolute deviation (MAD)
- ✓ Annotated box plot for outlier detection

D. R Statistical Code and AI Prompts

This section provides complete, reproducible R code for performing descriptive and univariate statistical analyses. The AI used was Claude.ai (Anthropic, 2026). Each code block represents an AI prompt that can be directly executed in R or R Studio. First of all, the dataset of snapping count observations for 94 tree species from Korup National Park (Table 1) should be copied to the AI prompt.

Table 1 Dataset of Snapping Count Observations for some Tree Species from Korup National Park

| Species | Snapping |
|--|----------|
| <i>Angylocalyx oligophyllus</i> (Bak.) Bak F. | 2 |
| <i>Annickia chrolanta</i> (Oliv.) Seten & P. J. Maas | 1 |
| <i>Anthonotha macrophylla</i> P. Beauv. | 1 |
| <i>Araliopsis soyauxii</i> Engl | 0 |
| <i>Aulocalyx talbotia</i> (Wernham) Keay | 1 |
| <i>Baphia capparidifolia</i> Bak. | 2 |
| <i>Barteria fistulosa</i> Mast. | 0 |
| <i>Berlinia auriculata</i> Benth. | 1 |
| <i>Calpocalyx dinklagei</i> Harms | 1 |
| <i>Carapa parviflora</i> Harms | 0 |

The following prompts should be typed below the pasted dataset, with respect to the steps in the analytical framework.

➤ *Create Vectors*

• *AI Prompt 1*

Act like R Statistics; for the data above, create the vectors and data frame.

✓ *Response*

▪ *Data Setup: Species and Snapping Frequency*

```
species <- c("Angylocalyx oligophyllus", "Annickia chrolanta", "Anthonotha macrophylla", "Araliopsis soyauxii", "Aulocalyx talbotia", "Baphia capparidifolia", "Barteria fistulosa", "Beilschmiedia sp.", "Berlinia auriculata", "Calpocalyx dinklagei", "Carapa parviflora", "Chytranthus talbotii", "Coelocaryon preussii", "Cola chlamydantha", "Cola acuminata", "Cola cauliflora", "Cola lateritia", "Cola praeacuta", "Cola rostrata", "Cola semecarpophylla", "Cola sp", "Crotonogynopsis sp", "Dacryodes edulis", "Dacryodes klaineana", "Dasylepis blackii", "Desbordesia glaucescens", "Dialium pachyphyllum", "Dialium sp", "Dichaetanthera africana", "Dichostema glaucascens", "Diogoia zenkeri", "Diospyros bipindensis", "Diospyros gabunensis", "Diospyros ituriensis", "Diospyros Zenkeri", "Drypetes principum", "Drypetes staudtii", "Endodesmia calophylloides", "Englerophytum sp", "Erisma delphus exsul", "Erythrophleum ivorense", "Garcinia conrauana", "Gilbertiodendron demonstrans", "Homalium africanum", "Homalium letestui", "Homalium longistylum", "Hymenostegia afzelii", "Hymenostegia bakeriana", "Hypodaphnis zenkeri", "Irvingia gabonensis", "Klaineanthus gaboniae", "Lecomtedoxa klaineana", "Lepidobotrys staudtii", "Lophira alata", "Maesobotrya dusenii", "Maesopsis eminii", "Mammea Africana", "Memecylon sp.", "Octoknema affinis", "Oubanguia alata", "Pauridiantha floribunda", "Pausinystelia macroceras", "Pentadesma butryacea", "Phyllobotryon spathulatum", "Poga oleosa", "Protomegabaria stapfiana", "Pycnocomma macrophylla", "Rauvolfia caffra", "Rhaptopetalum sp.", "Rinorea cf_woermanniana", "Rinorea gabunensis", "Rinorea lepidobotrys", "Scottellia klaineana", "Staudtia gabonensis", "Soyauxia talbotii", "Strephonema pseudocola", "Strombosia pustulata", "Strombosia scheffleri", "Strombosia sp", "Strombosiopsis tetrandra", "Strychnos sp", "Symphonia globulifera", "Synsepalum stipulatum", "Tabernaemontana
```

```
brachyantha", "Tabernaemontana crassa", "Talbotiella eketensis", "Tapura africana", "Trichoscypha sp", "Uapaca staudtii", "Uvariopsis bakeriana", "Vitex sp", "Warnackea Membranifolia", "Xylopia acutiflora", "Xylopia aethiopica")
```

```
SNAPPING <- c(2, 1, 1, 0, 1, 2, 0, 2, 1, 1, 0, 1, 1, 1, 2, 0, 2, 0, 2, 0, 1, 0, 2, 1, 1, 1, 0, 0, 0, 5, 2, 0, 3, 2, 0, 0, 0, 0, 2, 0, 0, 3, 0, 2, 1, 3, 3, 2, 2, 1, 3, 2, 1, 0, 1, 1, 1, 0, 0, 16, 0, 1, 1, 1, 1, 8, 1, 0, 0, 1, 1, 2, 3, 1, 2, 1, 1, 3, 0, 1, 1, 0, 2, 1, 1, 0, 1, 1, 1, 2, 1, 1)
```

Create a data frame

```
df <- data.frame(Species = species, Snapping = snapping)
cat("Data loaded successfully!\n")
cat("Number of species:", nrow(df), "\n")
cat("Snapping range:", min(snapping), "to", max(snapping), "\n")
```

➤ *Stage 1: Exploratory Data Analysis (EDA)*

• *AI Prompt 2*

Calculate Exploratory Data Analysis (EDA), following the points below:

- ✓ Data structure examination and summary statistics
- ✓ Calculation of measures of central tendency (mean, median, mode)
- ✓ Calculation of measures of dispersion (variance, standard deviation, IQR, MAD)
- ✓ Calculation of measures of shape (skewness, kurtosis)

• *Frequency Distribution Analysis*

- ✓ Histogram with superimposed normal curve
- ✓ Bar chart of frequency distribution
- ✓ Box plot with individual data points

• *Response*

Stage 1: Exploratory Data Analysis (EDA)

```
# --- 1.1 Data Structure & Overview ---
cat("--- Data Structure ---\n")
str(df)
cat("\n--- First 6 Rows ---\n")
print(head(df))
cat("\n--- Last 6 Rows ---\n")
print(tail(df))
```

```

# --- 1.2 Summary Statistics ---
cat("\n--- Summary Statistics ---\n")
print(summary(snapping))

# --- 1.3 Measures of Central Tendency ---
cat("\n--- Measures of Central Tendency ---\n")

mean_val <- mean(snapping)
median_val <- median(snapping)

# Mode function (base R has no built-in mode for numeric)
get_mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
mode_val <- get_mode(snapping)

cat("Mean :", round(mean_val, 4), "\n")
cat("Median :", median_val, "\n")
cat("Mode :", mode_val, "\n")

# --- 1.4 Measures of Dispersion ---
cat("\n--- Measures of Dispersion ---\n")

var_val <- var(snapping)
sd_val <- sd(snapping)
iqr_val <- IQR(snapping)
range_val <- range(snapping)
mad_val <- mad(snapping, constant = 1) # MAD without
scaling constant

cat("Variance :", round(var_val, 4), "\n")
cat("Standard Deviation :", round(sd_val, 4), "\n")
cat("IQR :", iqr_val, "\n")
cat("Range :", range_val[1], "to", range_val[2], "\n")
cat("MAD (unscaled) :", round(mad_val, 4), "\n")

# --- 1.5 Measures of Shape ---
cat("\n--- Measures of Shape ---\n")

# Install moments package if not available
if (!require(moments)) install.packages("moments",
dependencies = TRUE)
library(moments)

skew_val <- skewness(snapping)
kurt_val <- kurtosis(snapping)

cat("Skewness :", round(skew_val, 4),
ifelse(skew_val > 0, " → Positive (right) skew", " →
Negative (left) skew"), "\n")
cat("Kurtosis :", round(kurt_val, 4),
ifelse(kurt_val > 3, " → Leptokurtic (heavy tails)",
ifelse(kurt_val < 3, " → Platykurtic (light tails)", " →
Mesokurtic (normal)")), "\n")

# --- 1.6 Frequency Distribution ---
cat("\n--- Frequency Distribution Table ---\n")

freq_table <- as.data.frame(table(Snapping = snapping))

freq_table$RelFreq <- round(freq_table$Freq /
sum(freq_table$Freq), 4)
freq_table$CumFreq <- cumsum(freq_table$Freq)
freq_table$CumRelFreq <-
round(cumsum(freq_table$RelFreq), 4)
colnames(freq_table) <- c("Snapping_Count", "Frequency",
"Relative_Freq",
"Cumulative_Freq", "Cumulative_Rel_Freq")
print(freq_table)

# Set up a 2x3 multi-panel layout
par(mfrow = c(2, 3), mar = c(4, 4, 3, 1), oma = c(0, 0, 2, 0))

# --- 1.7 Histogram with Superimposed Normal Curve ---
hist(snapping,
breaks = 10,
freq = FALSE,
col = "steelblue",
border = "white",
main = "Histogram + Normal Curve",
xlab = "Snapping Frequency",
ylab = "Density",
las = 1)

# Superimpose normal curve
x_seq <- seq(min(snapping) - 1, max(snapping) + 1,
length.out = 200)
lines(x_seq,
dnorm(x_seq, mean = mean(snapping), sd = sd(snapping)),
col = "red", lwd = 2.5)
legend("topright", legend = "Normal curve", col = "red", lwd
= 2, cex = 0.8)

# --- 1.8 Bar Chart of Frequency Distribution ---
freq_counts <- table(snapping)
barplot(freq_counts,
col = "coral",
border = "white",
main = "Frequency Distribution Bar Chart",
xlab = "Snapping Count",
ylab = "Number of Species",
las = 1)

# Add count labels on top of bars
text(x = barplot(freq_counts, plot = FALSE),
y = freq_counts + 0.2,
labels = freq_counts,
cex = 0.8,
font = 2)

# Reset layout
mtext("Stage 2: Data Visualization - Snapping Frequency",
outer = TRUE, cex = 1.1, font = 2)
par(mfrow = c(1, 1))

# --- 1.9 Box Plot with Individual Data Points ---
boxplot(snapping,
col = "lightgreen",
border = "darkgreen",
main = "Box Plot with Data Points",
ylab = "Snapping Frequency",

```

```
outline = FALSE, # suppress default outlier points
las = 1)
```

```
# Jitter individual points
```

```
set.seed(123)
stripchart(snapping,
method = "jitter",
jitter = 0.15,
add = TRUE,
vertical = TRUE,
pch = 16,
col = adjustcolor("darkblue", alpha.f = 0.5),
cex = 0.8)
```

➤ Stage 2: Normality Testing

• AI Prompt

Perform multiple normality tests, following the points below:

- ✓ Shapiro-Wilk test: Most powerful for small to moderate sample sizes
- ✓ Anderson-Darling test: Gives more weight to the tails of the distribution
- ✓ Kolmogorov-Smirnov test: General goodness-of-fit test
- ✓ D'Agostino-Pearson test: Based on skewness and kurtosis
- ✓ Quantile-quantile (Q-Q) plot for normality assessment
- ✓ Kernel density plot

• Response

Stage 2: Normality Testing

```
# Load packages - install only if truly missing, with error handling
```

```
for (pkg in c("nortest", "moments")) {
if (!requireNamespace(pkg, quietly = TRUE)) {
tryCatch(
install.packages(pkg, repos = "https://cloud.r-project.org",
dependencies = TRUE),
error = function(e) message("Could not install ", pkg, ": ",
e$message)
)
}
suppressPackageStartupMessages(library(pkg, character.only
= TRUE))
}
```

```
# --- 2.1 Shapiro-Wilk Test ---
```

```
cat("--- Shapiro-Wilk Test ---\n")
sw_test <- shapiro.test(snapping)
print(sw_test)
cat("Interpretation: W =", round(sw_test$statistic, 4),
"| p-value =", round(sw_test$p.value, 6), "\n")
cat(ifelse(sw_test$p.value < 0.05,
"-> REJECT H0: Data is NOT normally distributed (p <
0.05)\n",
"-> FAIL TO REJECT H0: Data may be normally distributed
(p >= 0.05)\n"))
```

```
# --- 2.2 Anderson-Darling Test ---
```

```
cat("\n--- Anderson-Darling Test ---\n")
ad_test <- ad.test(snapping)
print(ad_test)
cat("Interpretation: A =", round(ad_test$statistic, 4),
"| p-value =", round(ad_test$p.value, 6), "\n")
cat(ifelse(ad_test$p.value < 0.05,
"-> REJECT H0: Data is NOT normally distributed (p <
0.05)\n",
"-> FAIL TO REJECT H0: Data may be normally distributed
(p >= 0.05)\n"))
```

```
# --- 2.3 Kolmogorov-Smirnov Test ---
```

```
cat("\n--- Kolmogorov-Smirnov Test (vs. Normal) ---\n")
ks_test <- ks.test(snapping, "pnorm",
mean = mean(snapping),
sd = sd(snapping))
print(ks_test)
cat("Interpretation: D =", round(ks_test$statistic, 4),
"| p-value =", round(ks_test$p.value, 6), "\n")
cat(ifelse(ks_test$p.value < 0.05,
"-> REJECT H0: Data is NOT normally distributed (p <
0.05)\n",
"-> FAIL TO REJECT H0: Data may be normally distributed
(p >= 0.05)\n"))
```

```
# --- 2.4 D'Agostino-Pearson Test ---
```

```
cat("\n--- D'Agostino-Pearson Test (Skewness + Kurtosis) ---
\n")
dp_test <- agostino.test(snapping)
print(dp_test)
cat("Interpretation: p-value =", round(dp_test$p.value, 6),
"\n")
cat(ifelse(dp_test$p.value < 0.05,
"-> REJECT H0: Significant departure from normality (p <
0.05)\n",
"-> FAIL TO REJECT H0: No significant departure from
normality (p >= 0.05)\n"))
```

```
# --- Summary Table ---
```

```
cat("\n--- Normality Tests Summary ---\n")
norm_summary <- data.frame(
Test = c("Shapiro-Wilk", "Anderson-Darling",
"Kolmogorov-Smirnov", "D'Agostino-Pearson"),
Statistic = round(c(sw_test$statistic, ad_test$statistic,
ks_test$statistic, dp_test$statistic), 4),
P_Value = round(c(sw_test$p.value, ad_test$p.value,
ks_test$p.value, dp_test$p.value), 6),
Decision = ifelse(c(sw_test$p.value, ad_test$p.value,
ks_test$p.value, dp_test$p.value) < 0.05,
"Non-Normal", "Normal")
)
print(norm_summary)
```

```
# Set up a 2x3 multi-panel layout
```

```
par(mfrow = c(2, 3), mar = c(4, 4, 3, 1), oma = c(0, 0, 2, 0))
```

```
# --- 2.5 Q-Q Plot for Normality ---
```

```
qqnorm(snapping,
main = "Q-Q Plot (Normality Check)",
pch = 16,
```

```
col = "darkred",
cex = 0.8,
las = 1)
qqline(snapping, col = "blue", lwd = 2, lty = 2)

# --- 2.6 Kernel Density Plot ---
dens <- density(snapping, bw = "nrd0")
plot(dens,
main = "Kernel Density Plot",
xlab = "Snapping Frequency",
ylab = "Density",
col = "purple",
lwd = 2.5,
las = 1)
polygon(dens, col = adjustcolor("purple", alpha.f = 0.2),
border = NA)
abline(v = mean(snapping), col = "red", lty = 2, lwd = 1.5)
abline(v = median(snapping), col = "orange", lty = 2, lwd =
1.5)
legend("topright",
legend = c("Mean", "Median"),
col = c("red", "orange"),
lty = 2, lwd = 1.5, cex = 0.75)
```

➤ Stage 3: Hypothesis Testing

• AI Prompt

Conduct parametric and non-parametric tests, following the points below:

Parametric Methods (assuming normality):

- ✓ One-sample t-test (two hypotheses: $\mu = 0$ and $\mu = 1$)
- ✓ Z-test (large sample approximation)
- ✓ Confidence intervals for the population mean
- ✓ Non-Parametric Methods (distribution-free):
- ✓ Wilcoxon signed-rank test
- ✓ Sign test
- ✓ Mann-Whitney U test
- ✓ Bootstrap confidence intervals for the median

• Response

Stage 3: Hypothesis Testing

```
# --- Load BSDA safely; fall back to manual sign test if
unavailable ---
bsdavaavailable <- requireNamespace("BSDA", quietly =
TRUE)
if (!bsdavaavailable) {
tryCatch(
{ install.packages("BSDA", repos = "https://cloud.r-
project.org", dependencies = TRUE)
bsdavaavailable <- requireNamespace("BSDA", quietly =
TRUE) },
error = function(e) message("BSDA not available. Manual
sign test will be used instead."),
warning = function(w) message("BSDA not available.
Manual sign test will be used instead.")
)
}
```

```
if (bsdavaavailable)
suppressPackageStartupMessages(library(BSDA))

n <- length(snapping)
x_bar <- mean(snapping)
s <- sd(snapping)
se <- s / sqrt(n)

# ---- PARAMETRIC METHODS ----
cat("=====  
PARAMETRIC METHODS =====\n\n")

# --- 3.1 One-Sample t-test: H0:  $\mu = 0$  ---
cat("--- One-Sample t-test: H0:  $\mu = 0$  ---\n")
t_test_0 <- t.test(snapping, mu = 0, alternative = "two.sided",
conf.level = 0.95)
print(t_test_0)
cat("Decision:", ifelse(t_test_0$p.value < 0.05,
"REJECT H0 - Mean significantly differs from 0\n",
"FAIL TO REJECT H0 - No significant difference from
0\n"))

# --- 3.2 One-Sample t-test: H0:  $\mu = 1$  ---
cat("\n--- One-Sample t-test: H0:  $\mu = 1$  ---\n")
t_test_1 <- t.test(snapping, mu = 1, alternative = "two.sided",
conf.level = 0.95)
print(t_test_1)
cat("Decision:", ifelse(t_test_1$p.value < 0.05,
"REJECT H0 - Mean significantly differs from 1\n",
"FAIL TO REJECT H0 - No significant difference from
1\n"))

# --- 3.3 Z-Test (large sample approximation) ---
cat("\n--- Z-Test (Large Sample Approximation) ---\n")

z_test_manual <- function(x, mu0) {
n_x <- length(x)
z_stat <- (mean(x) - mu0) / (sd(x) / sqrt(n_x))
p_val <- 2 * pnorm(-abs(z_stat))
ci_low <- mean(x) - 1.96 * (sd(x) / sqrt(n_x))
ci_up <- mean(x) + 1.96 * (sd(x) / sqrt(n_x))
cat("Z-statistic :", round(z_stat, 4), "\n")
cat("p-value :", round(p_val, 6), "\n")
cat("95% CI :", round(ci_low, 4), "to", round(ci_up, 4),
"\n")
cat("Decision :", ifelse(p_val < 0.05, "REJECT H0\n",
"FAIL TO REJECT H0\n"))
}

cat("H0:  $\mu = 0$ \n"); z_test_manual(snapping, mu0 = 0)
cat("\nH0:  $\mu = 1$ \n"); z_test_manual(snapping, mu0 = 1)

# --- 3.4 Confidence Intervals for Population Mean ---
cat("\n--- Confidence Intervals for Population Mean ---\n")
ci_90 <- t.test(snapping, conf.level = 0.90)$conf.int
ci_95 <- t.test(snapping, conf.level = 0.95)$conf.int
ci_99 <- t.test(snapping, conf.level = 0.99)$conf.int
cat("90% CI:", round(ci_90[1], 4), "to", round(ci_90[2], 4),
"\n")
cat("95% CI:", round(ci_95[1], 4), "to", round(ci_95[2], 4),
"\n")
```

```
cat("99% CI:", round(ci_99[1], 4), "to", round(ci_99[2], 4),
"\n")
```

```
# ---- NON-PARAMETRIC METHODS ----
```

```
cat("\n==== NON-PARAMETRIC METHODS =====\n\n")
```

```
# --- 3.5 Wilcoxon Signed-Rank Test: H0: median = 0 ---
```

```
cat("--- Wilcoxon Signed-Rank Test: H0: median = 0 ---\n")
wilcox_0 <- wilcox.test(snapping, mu = 0, alternative =
"two.sided", conf.int = TRUE)
print(wilcox_0)
cat("Decision:", ifelse(wilcox_0$p.value < 0.05,
"REJECT H0 - Median significantly differs from 0\n",
"FAIL TO REJECT H0\n"))
```

```
# --- 3.6 Wilcoxon Signed-Rank Test: H0: median = 1 ---
```

```
cat("\n--- Wilcoxon Signed-Rank Test: H0: median = 1 ---
\n")
wilcox_1 <- wilcox.test(snapping, mu = 1, alternative =
"two.sided", conf.int = TRUE)
print(wilcox_1)
cat("Decision:", ifelse(wilcox_1$p.value < 0.05,
"REJECT H0 - Median significantly differs from 1\n",
"FAIL TO REJECT H0\n"))
```

```
# --- 3.7 Sign Test: H0: median = 1 ---
```

```
cat("\n--- Sign Test: H0: median = 1 ---\n")
```

```
if (bsda_available) {
# Use BSDA if available
sign_result <- SIGN.test(snapping, md = 1, alternative =
"two.sided", conf.level = 0.95)
print(sign_result)
```

```
} else {
```

```
# Manual sign test - no packages needed
```

```
md0 <- 1
pos <- sum(snapping > md0) # values above null
median
neg <- sum(snapping < md0) # values below null
median
ties <- sum(snapping == md0) # excluded from test
n_eff <- pos + neg # effective sample size
# Two-tailed p-value using binomial distribution
p_sign <- 2 * min(
pbinom(min(pos, neg), size = n_eff, prob = 0.5),
1 - pbinom(min(pos, neg) - 1, size = n_eff, prob = 0.5)
)
```

```
# Exact binomial 95% CI for the median (Clopper-Pearson)
```

```
ci_sign <- binom.test(pos, n_eff, p = 0.5)$conf.int
```

```
cat("Sign Test (manual, base R)\n")
cat("Null median (md) :", md0, "\n")
cat("Observations above md :", pos, "\n")
cat("Observations below md :", neg, "\n")
cat("Ties (excluded) :", ties, "\n")
cat("Effective n :", n_eff, "\n")
cat("p-value (two-tailed) :", round(p_sign, 6), "\n")
cat("95% CI (Clopper-Pearson): [", round(ci_sign[1], 4), ", ",
round(ci_sign[2], 4), "]\n")
cat("Decision:", ifelse(p_sign < 0.05,
```

```
"REJECT H0 - Median significantly differs from 1\n",
"FAIL TO REJECT H0 - No significant difference from
1\n"))
}
```

```
# --- 3.8 Mann-Whitney U Test (vs. uniform baseline of 1) ---
```

```
cat("\n--- Mann-Whitney U Test (snapping vs. uniform
baseline of 1) ---\n")
set.seed(42)
null_population <- rep(1, n)
mw_test <- wilcox.test(snapping, null_population, alternative
= "two.sided", exact = FALSE)
print(mw_test)
cat("Decision:", ifelse(mw_test$p.value < 0.05,
"REJECT H0 - Distribution differs from baseline\n",
"FAIL TO REJECT H0\n"))
```

```
# --- 3.9 Bootstrap Confidence Interval for the Median ---
```

```
cat("\n--- Bootstrap 95% CI for the Median (B = 10,000) ---
\n")
set.seed(123)
B <- 10000
boot_medians <- replicate(B, median(sample(snapping, n,
replace = TRUE)))
boot_ci <- quantile(boot_medians, probs = c(0.025,
0.975))
```

```
cat("Observed Median :", median(snapping), "\n")
cat("Bootstrap 95% CI :", round(boot_ci[1], 4), "to",
round(boot_ci[2], 4), "\n")
cat("Bootstrap SE of Median :", round(sd(boot_medians), 4),
"\n")
```

```
hist(boot_medians,
breaks = 30,
col = "lightblue",
border = "white",
main = "Bootstrap Distribution of Median",
xlab = "Bootstrapped Median",
las = 1)
abline(v = boot_ci, col = "red", lty = 2, lwd = 2)
abline(v = median(snapping), col = "darkblue", lwd = 2)
legend("topright",
legend = c("Observed Median", "95% CI"),
col = c("darkblue", "red"),
lty = c(1, 2), lwd = 2, cex = 0.85)
```

➤ Stage 4: Outlier Detection

• AI Prompt

Detect outliers using multiple methods, following the points below:

- ✓ IQR method ($1.5 \times \text{IQR}$ rule)
- ✓ Z-score method ($|z| > 3$)
- ✓ Modified Z-score using median absolute deviation (MAD)
- ✓ Annotated box plot for outlier detection

- *Response*

Stage 4: Outlier Detection

```
# --- 4.1 IQR Method (1.5 × IQR Rule) ---
cat("--- Method 1: IQR Rule (1.5 × IQR) ---\n")
Q1 <- quantile(snapping, 0.25)
Q3 <- quantile(snapping, 0.75)
IQR_val <- IQR(snapping)
lower_iqr <- Q1 - 1.5 * IQR_val
upper_iqr <- Q3 + 1.5 * IQR_val

iqr_outliers <- df[snapping < lower_iqr | snapping >
upper_iqr, ]
cat("Q1:", Q1, "| Q3:", Q3, "| IQR:", IQR_val, "\n")
cat("Lower fence:", lower_iqr, "| Upper fence:", upper_iqr,
"\n")
cat("Number of outliers:", nrow(iqr_outliers), "\n")
if (nrow(iqr_outliers) > 0) print(iqr_outliers)

# --- 4.2 Z-Score Method (|z| > 3) ---
cat("\n--- Method 2: Z-Score Method (|z| > 3) ---\n")
z_scores <- (snapping - mean(snapping)) / sd(snapping)
z_outliers <- df[abs(z_scores) > 3, ]
z_outlier_z <- z_scores[abs(z_scores) > 3]

cat("Number of outliers (|z| > 3):", nrow(z_outliers), "\n")
if (nrow(z_outliers) > 0) {
z_result <- data.frame(z_outliers, Z_Score =
round(z_outlier_z, 4))
print(z_result)
}

# --- 4.3 Modified Z-Score Using MAD ---
cat("\n--- Method 3: Modified Z-Score (MAD-based,
threshold = 3.5) ---\n")
med_val <- median(snapping)
mad_raw <- mad(snapping, constant = 1) # unscaled
MAD
mod_z <- 0.6745 * (snapping - med_val) / mad_raw
mod_outliers <- df[abs(mod_z) > 3.5, ]
mod_z_vals <- mod_z[abs(mod_z) > 3.5]

cat("Median:", med_val, "| MAD (unscaled):",
round(mad_raw, 4), "\n")
cat("Number of outliers (|Mod Z| > 3.5):",
nrow(mod_outliers), "\n")
if (nrow(mod_outliers) > 0) {
mod_result <- data.frame(mod_outliers, Modified_Z =
round(mod_z_vals, 4))
print(mod_result)
}

# --- 4.4 Outlier Summary Comparison ---
cat("\n--- Outlier Detection Summary ---\n")
outlier_summary <- data.frame(
Method = c("IQR (1.5 × IQR)", "Z-Score (|z|>3)",
"Modified Z-Score (|mz|>3.5)"),
N_Outliers = c(nrow(iqr_outliers), nrow(z_outliers),
nrow(mod_outliers)),
Species_Found = c(

```

```
ifelse(nrow(iqr_outliers) > 0, paste(iqr_outliers$Species,
collapse = "; "), "None"),
ifelse(nrow(z_outliers) > 0, paste(z_outliers$Species,
collapse = "; "), "None"),
ifelse(nrow(mod_outliers) > 0, paste(mod_outliers$Species,
collapse = "; "), "None")
)
)
print(outlier_summary)
```

```
# --- 4.5 Visualization: Annotated Box Plot with Outlier
Flags ---
par(mar = c(8, 4, 4, 2))
boxplot(snapping,
col = "lightyellow",
border = "darkgoldenrod",
main = "Outlier Detection - Annotated Box Plot",
ylab = "Snapping Frequency",
outline = TRUE,
las = 1)
```

```
# Add colour-coded points by outlier method
points(rep(1, n), snapping,
pch = 16, col = adjustcolor("gray50", 0.4), cex = 0.9)
```

```
# Highlight IQR outliers in red
if (nrow(iqr_outliers) > 0) {
points(rep(1, nrow(iqr_outliers)), iqr_outliers$Snapping,
pch = 17, col = "red", cex = 1.5)
text(rep(1.12, nrow(iqr_outliers)), iqr_outliers$Snapping,
labels = iqr_outliers$Species,
cex = 0.65, col = "red", adj = 0)
}
```

```
# Add fence lines
abline(h = upper_iqr, col = "red", lty = 2, lwd = 1.5)
abline(h = lower_iqr, col = "orange", lty = 2, lwd = 1.5)
legend("topright",
legend = c("All points", "IQR Outlier", "Upper fence",
"Lower fence"),
col = c("gray50", "red", "red", "orange"),
pch = c(16, 17, NA, NA),
lty = c(NA, NA, 2, 2),
lwd = c(NA, NA, 1.5, 1.5),
cex = 0.8)
par(mar = c(5, 4, 4, 2)) # reset margins
```

E. Comparison of Vibe Analysis and Manual Analysis

To evaluate the reliability and practical utility of AI-assisted statistical analysis (vibe analysis) relative to conventional analyst-coded analysis (manual analysis), both approaches were applied in parallel to the same snapping count dataset (n = 94 tree species, Korup National Park) using an identical four-stage analytical framework. The comparison was structured around five dimensions: numerical agreement, assumption-checking completeness, method-selection alignment, ecological interpretive alignment, and process efficiency. Each dimension corresponded to a pre-defined evaluation criterion assessed after both analyses were independently completed.

➤ Numerical Agreement

To assess numerical agreement between the two approaches, each of the 16 statistical outputs reported in both analyses, comprising point estimates (mean, median, standard deviation, skewness, kurtosis), test statistics and associated p-values (Shapiro-Wilk W, Anderson-Darling A, Kolmogorov-Smirnov D, one-sample t-test t, Wilcoxon signed-rank V, sign test p), outlier counts and identified values (IQR method, Z-score method, modified Z-score method), the designated primary inferential method, and the overall statistical conclusion were compared directly between the vibe analysis output and the manual analysis output. Agreement was defined as exact equivalence to three or more decimal places. Any discrepancy, however small, was to be recorded and investigated as a potential indicator of LLM numerical hallucination. To safeguard against hallucination, all statistics reported by the vibe analysis were verified by independently executing the AI-generated R code in R and comparing the console output to the values stated in the LLM's narrative response; only code-executed values were used in the comparison.

➤ Assumption-Checking and Method Selection

Assumption-checking procedures were conducted identically under both approaches prior to any inferential testing. Normality was assessed using four formal statistical tests: the Shapiro-Wilk test, the Anderson-Darling test, the Kolmogorov-Smirnov test, and the D'Agostino-Pearson test. Graphical diagnostics, comprising a histogram overlaid with a fitted normal distribution curve, a quantile-quantile (Q-Q) plot, a kernel density estimate, and a box plot were produced and examined alongside the formal test results. Outliers were identified using three independent methods: the interquartile range (IQR) fence method, the Z-score method (threshold $|Z| > 3$), and the modified Z-score method based on the median absolute deviation (MAD), using the threshold of 3.5 recommended by Iglewicz and Hoaglin (1993).

Method selection was treated as a decision point conditional on the outcomes of the assumption-checking phase rather than pre-assigned. Under the manual analysis, the analyst reviewed all normality test statistics and graphical diagnostics before formally designating non-parametric tests (Wilcoxon signed-rank test, sign test, bootstrap confidence intervals) as the primary inferential tools, retaining the one-sample t-test for comparative purposes only, with explicit caveats regarding normality violation. Under the vibe analysis, the LLM was not instructed which tests to select; method designation was left entirely to the AI's autonomous assessment of the diagnostic outputs it generated. The method ultimately selected by each approach, and the reasoning provided, were recorded and compared to determine whether the AI replicated the analyst's pre-specified decision protocol without direct instruction.

➤ Decision Process Comparison

To characterise how each approach arrived at its analytical conclusions, the decision pathway was traced across five key decision points: (i) recognition of non-normality, (ii) selection of the primary inferential test, (iii) retention of parametric tests for comparative purposes, (iv) identification

of the outlier detection method, and (v) production of ecological interpretation. For each decision point, the mechanism by which each approach reached its conclusion was documented, including the evidence consulted, the sequence of steps taken, and whether the process was iterative or resolved within a single analytical step. This comparison was designed to expose structural differences in the reasoning process between human analyst judgment and LLM-autonomous decision-making, even where the final conclusions were identical.

➤ Interpretive Alignment

Forestry interpretation produced by each approach was compared qualitatively following completion of the statistical analyses. Both approaches were evaluated on three criteria: (i) whether outlier species were correctly identified and characterised as biologically distinct; (ii) whether the disproportionate influence of extreme observations on distributional shape (skewness, heavy right tail) was recognised and contextualised; and (iii) whether the overall statistical conclusion regarding the population median was ecologically coherent and consistent with the inferential test outcomes. Under the manual analysis, forestry framing was provided by the analyst drawing on field expertise and published knowledge of Korup National Park tree species. Under the vibe analysis, the LLM was not prompted to provide ecological interpretation; any ecological narrative produced was generated autonomously from the statistical outputs alone. The depth and accuracy of each approach's forestry contextualisation were assessed, with particular attention to whether the AI independently linked statistical anomalies, specifically the extreme outlier counts, to biologically meaningful explanations without analyst guidance.

➤ Efficiency, Risk, And Validation

Process efficiency was assessed by comparing the number of analytical steps, coding cycles, and iterative decisions required to complete all eight analytical stages under each approach. The vibe analysis was conducted within prompt-response sessions, with R code, graphical diagnostics, and interpretive narrative generated concurrently by the LLM. The manual analysis required sequential coding, individual debugging, and stage-by-stage interpretation. Efficiency was therefore characterised in terms of relative time investment and analytical workflow complexity rather than absolute clock time, as the latter was not formally recorded.

The risk of LLM hallucination in quantitative reporting was addressed through a mandatory validation step: every numerical value stated in the vibe analysis narrative was independently verified by executing the AI-generated R code in R and comparing the computed output to the reported value. An output was accepted as validated only upon exact agreement between the two. This validation step was applied to all 16 reported statistical outputs. Hallucination risk was further contextualised by acknowledging the analytical complexity of the present dataset; a single-variable, moderate-sample-size count dataset, relative to more complex forestry data structures that may expose LLM limitations not detectable in the current case.

III. RESULTS

This section presents the statistical analysis of snapping frequency across 94 tree species, aiming to characterise its distribution, central tendency, and variability. Both descriptive and inferential methods were applied to evaluate patterns and test key hypotheses regarding mean and median snapping rates. Given the highly skewed and non-normal nature of the data, a combination of parametric, non-parametric, and bootstrap approaches was employed. The results provide robust insights into the prevalence, distributional behaviour, and ecological significance of snapping events within the species assemblage.

A. Descriptive Statistics

Snapping frequency data were collected from 94 tree species. The dataset exhibited a wide range of values, spanning from 0 (no snapping events recorded) to a maximum of 16 snapping events for a single species.

As Table 2 illustrates, the mean snapping frequency was 1.33 events per species (SD = 1.96), while the median was

1.0, indicating that the central tendency measures are in close agreement for the bulk of the distribution. The mode was also 1, suggesting that single snapping events were the most commonly recorded observation across species. The standard deviation of 1.96 indicates moderate spread around the mean; however, this value is disproportionately influenced by extreme observations. The interquartile range (IQR = 2.0) and the median absolute deviation (MAD = 1.0) provide more robust measures of dispersion. The IQR indicates that the central 50% of species recorded between 0 and 2 snapping events, confirming that most species snapped infrequently. The distribution was characterised by extremely high positive skewness (5.057) and leptokurtic shape (36.32), both indicative of a heavily right-skewed, long-tailed distribution. Such values arise from the presence of a small number of species with unusually high snapping frequencies, most notably *Oubanguia alata* (16) and *Protomegabaria stapfiana* (8), which exert strong leverage on the parametric moments. These characteristics have direct implications for the choice of appropriate statistical tests, as detailed in subsequent sections.

Table 2 Descriptive Statistics for Snapping Frequency Across 94 Tree Species.

| Statistic | Value |
|---------------------------------|----------------------|
| Sample size (n) | 94 |
| Minimum | 0 |
| Maximum | 16 |
| Range | 16 |
| Mean (\bar{x}) | 1.3333 |
| Median | 1.0 |
| Mode | 1 |
| Variance (s^2) | 3.8333 |
| Standard Deviation (s) | 1.9579 |
| Interquartile Range (IQR) | 2.0 (Q1 = 0, Q3 = 2) |
| Median Absolute Deviation (MAD) | 1.0 |
| Skewness | 5.057 |
| Kurtosis (Pearson) | 36.320 |

The frequency distribution (Table 3) reveals that the majority of species fell in the lower snapping categories as 29.0% of species recorded no snapping events, 37.6% recorded exactly one event, and 18.3% recorded two events. Cumulatively, 84.9% of species recorded two or fewer

snapping events, confirming that infrequent snapping is the norm within this community. The three highest values (5, 8, and 16) each belonged to a single species, collectively representing only 3.2% of the sample yet exerting substantial influence on parametric summary statistics.

Table 3 Frequency Distribution of Snapping Frequency Values Across 94 Tree Species.

| Snapping Count | F | RF | CF | CRF |
|----------------|----|--------|----|--------|
| 0 | 27 | 0.2903 | 27 | 0.2903 |
| 1 | 35 | 0.3763 | 62 | 0.6667 |
| 2 | 17 | 0.1828 | 79 | 0.8495 |
| 3 | 8 | 0.0860 | 87 | 0.9355 |
| 5 | 1 | 0.0108 | 88 | 0.9462 |
| 8 | 1 | 0.0108 | 89 | 0.9570 |
| 16 | 1 | 0.0108 | 90 | 0.9677 |

F = Frequency, RF = Relative Frequency, CF = Cumulative Frequency, CRF = Cumulative Relative Frequency

The histogram with normal curve (Figure 1) reveals a strongly right-skewed, unimodal distribution. The superimposed normal curve, parameterised by the sample mean (1.33) and standard deviation (1.96), demonstrates a

severe departure from normality: the empirical distribution is concentrated at zero and one, with an extended right tail, in stark contrast to the symmetric bell-shaped expectation.

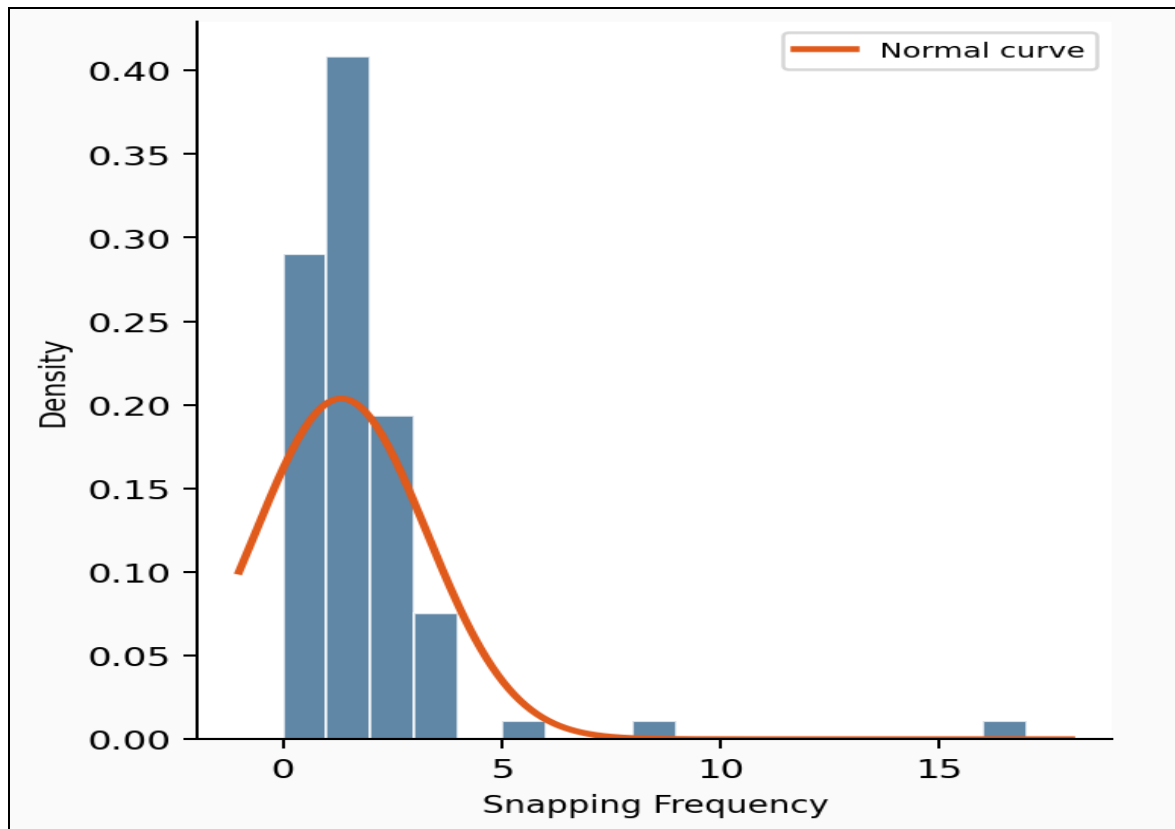


Fig 1 Histogram with Normal Curve of Snapping Frequency Across 94 Tree Species

The frequency distribution bar chart (Figure 2) confirms the discrete, count-based nature of the variable and the dominance of low-frequency snapping values. The two

categories 0 and 1 together account for over 66.7% of all observations, consistent with a zero-inflated or geometric-type distribution rather than a normal one.

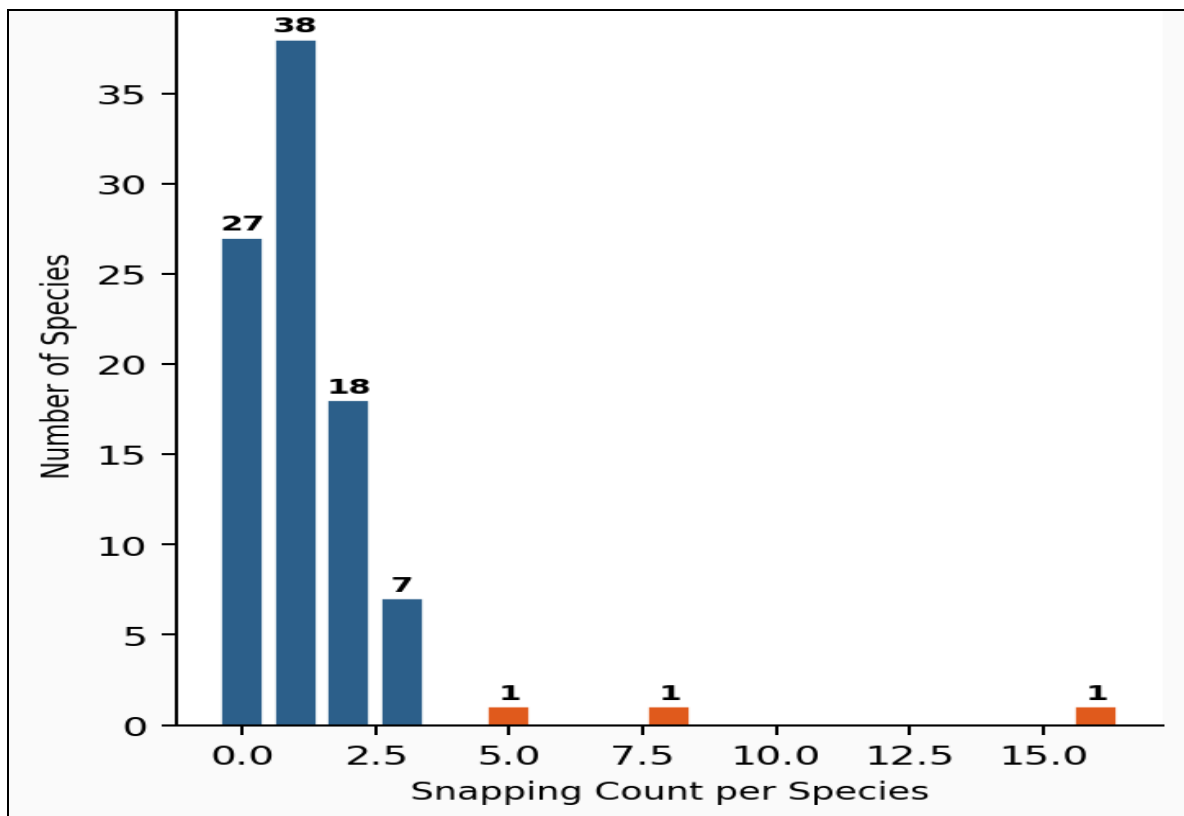


Fig 2 Frequency Distribution Bar Chart of Snapping Frequency Across 94 Tree Species

Box plot with individual data points (Figure 3) confirms that the interquartile range is narrow (IQR = 2), with the median (Q2 = 1) positioned at the lower end of the box, consistent with positive skew. Individual jittered data points

illustrate the discrete, count-based nature of the variable and highlight the high concentration of observations at values 0 and 1, as well as the presence of extreme high-value observations beyond the upper whisker.

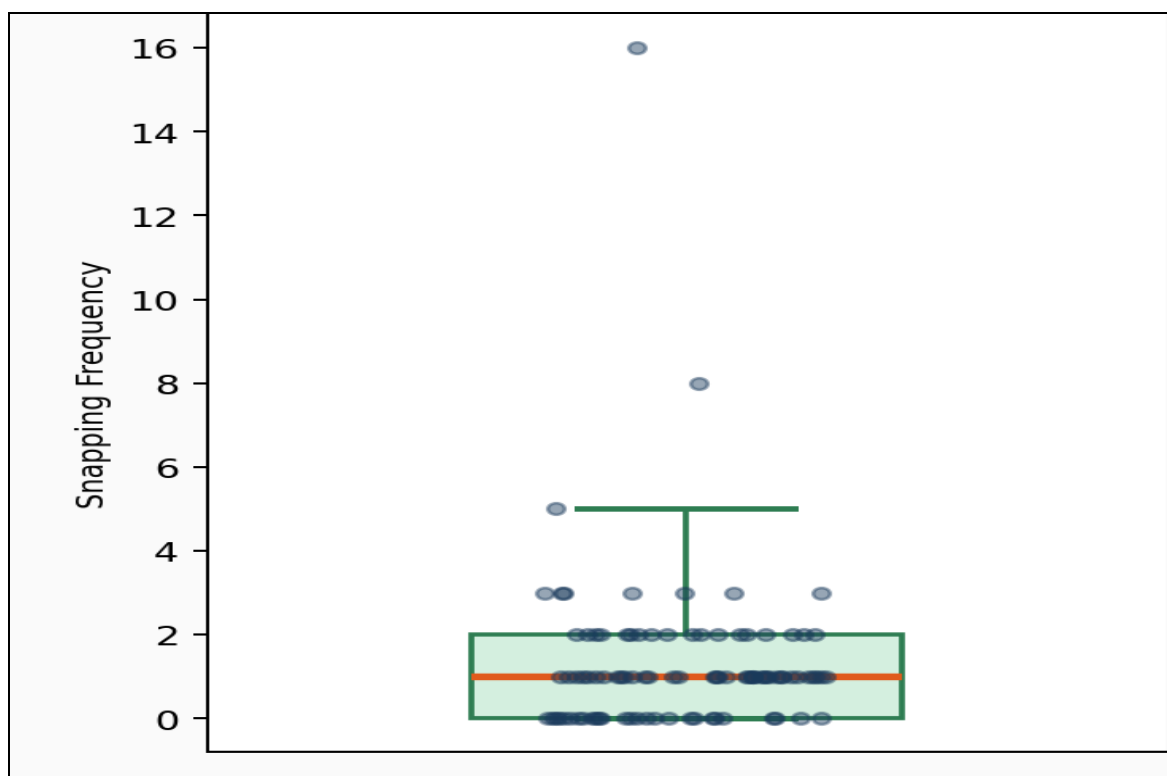


Fig 3 Box Plot with Individual Data Points of Snapping Frequency Across 94 Tree Species

B. Normality Testing

To formally assess whether snapping frequency follows a normal distribution, a prerequisite assumption for parametric hypothesis tests, four independent normality tests were applied.

All four normality tests rejected the null hypothesis of normality at the $\alpha = 0.05$ significance level, with all p-values falling well below 0.0001. Shapiro-Wilk test returned a statistic of $W = 0.5243$. A W value approaching 1.0 would indicate perfect normality; the observed value of 0.52 represents a severe departure, consistent with the high skewness and kurtosis of the distribution ($p < 0.0001$). The Anderson-Darling statistic ($A = 10.109$) substantially exceeded all critical values, including the most stringent

threshold at the 1% significance level (critical value = 1.026), providing overwhelming evidence against normality. This test is particularly sensitive to departures in the distribution tails, making it well-suited for detecting the extreme upper-tail behaviour observed here. One-sample Kolmogorov-Smirnov test comparing the empirical distribution to a normal distribution parameterised by the sample mean and standard deviation, returned $D = 0.2665$ ($p < 0.0001$). This large D statistic reflects substantial deviation between the observed cumulative distribution and the theoretical normal. D'Agostino-Pearson omnibus test confirms that both the asymmetry (skewness = 5.06) and the heavy-tailedness (kurtosis = 36.32) of the distribution depart significantly from normal expectations (Table 4).

Table 4 Normality Tests Applied to Snapping Frequency Data (n = 94). H_0 : Data are Normally Distributed

| Test | Test Statistic | p-value | Decision ($\alpha = 0.05$) |
|--------------------|----------------|------------|------------------------------|
| Shapiro-Wilk | $W = 0.5243$ | < 0.0001 | Reject H_0 |
| Anderson-Darling | $A = 10.109$ | < 0.001 | Reject H_0 |
| Kolmogorov-Smirnov | $D = 0.2665$ | < 0.0001 | Reject H_0 |
| D'Agostino-Pearson | $K^2 = 135.50$ | < 0.0001 | Reject H_0 |

Taken together, these four independent tests provide unequivocal evidence that snapping frequency is not normally distributed. Consequently, non-parametric methods and their results are given primary interpretive weight in hypothesis

testing, while parametric results are presented for completeness and comparative purposes.

Figure 4 shows the quantile-quantile (Q-Q) plot, a visual assessment of normality. If the data followed a normal

distribution, observed quantiles would align closely with the reference diagonal line. The pronounced S-shaped deviation with points curving sharply away from the line at both tails

confirms the non-normal character of the distribution, particularly in the upper tail where the two extreme outlier species cause substantial departure.

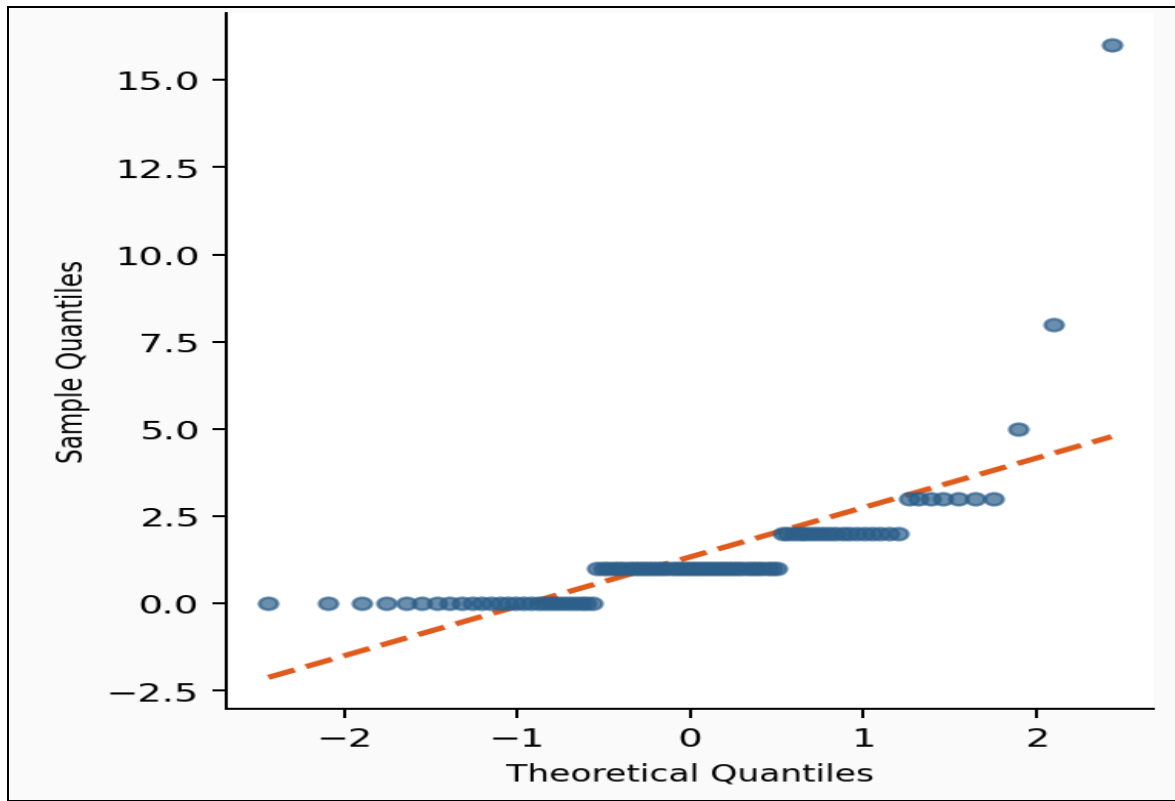


Fig 4 Quantile-Quantile (Q-Q) Plot of Snapping Frequency Across 94 Tree Species

Kernel density estimate plot (Figure 5) reveals the smoothed probability density function of snapping frequency. The distribution exhibits a sharp peak near 0 and 1, followed by a gradual monotone decrease toward higher values. The

separation between the mean (dashed red line, $\bar{x} = 1.33$) and median (dotted green line, $Md = 1.0$) further illustrates the influence of extreme values pulling the arithmetic mean rightward.

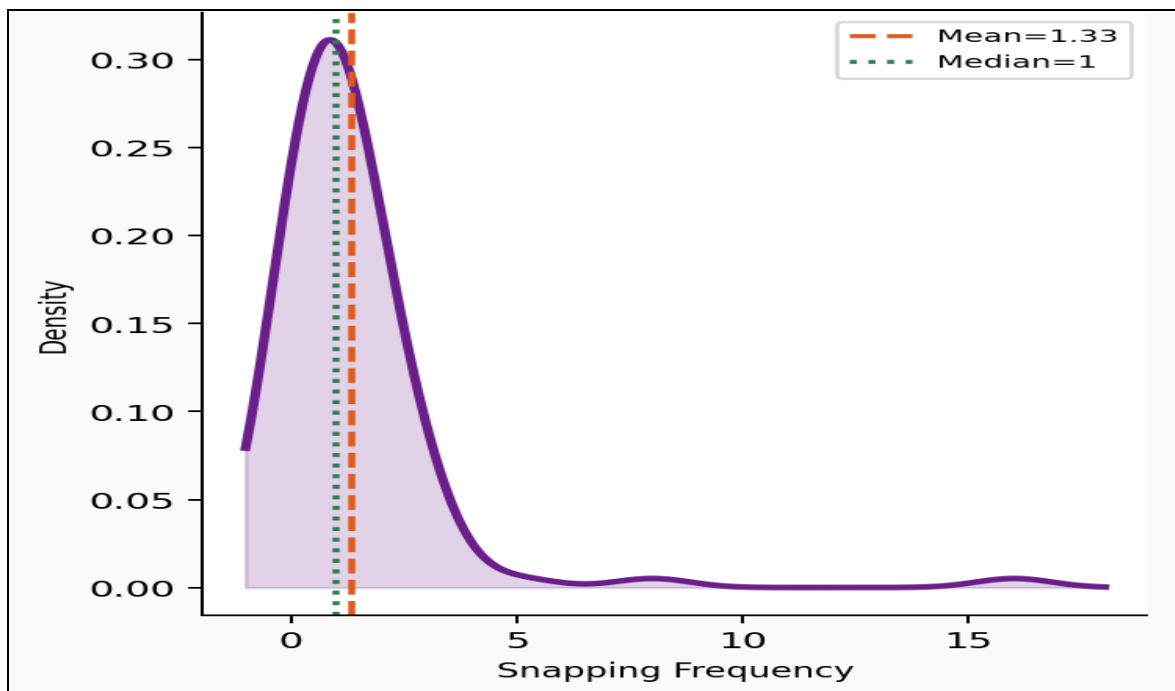


Fig 5 Kernel Density Estimate with Mean and Median Lines of Snapping Frequency Across 94 Tree Species

C. Hypothesis Testing

➤ Parametric Test Results

Two complementary hypotheses were tested: ($H_0: \mu = 0$) to determine whether snapping events were absent on average, and ($H_0: \mu = 1$) to assess whether the mean snapping frequency equalled one event per species. Both parametric and non-parametric approaches were employed given the confirmed non-normality of the data.

The one-sample t-test testing $H_0: \mu = 0$ was highly significant ($t(92) = 6.567, p < 0.0001$), providing strong evidence that the true population mean snapping frequency is significantly greater than zero. This confirms that snapping is a genuine biological phenomenon across the species assemblage, not a product of sampling error. In contrast, the test for $H_0: \mu = 1$ was non-significant ($t(92) = 1.642, p = 0.104$), indicating that the data are consistent with a population mean of one snapping event per species (Table 5).

Table 5 Parametric Hypothesis Test Results for Snapping Frequency. Degrees of Freedom for t-tests = 92.

| Test | H_0 | Statistic | p-value | Decision |
|-------------------|-----------|-----------------|--------------|----------------------|
| One-sample t-test | $\mu = 0$ | $t(92) = 6.567$ | < 0.0001 | Reject H_0 |
| One-sample t-test | $\mu = 1$ | $t(92) = 1.642$ | $p = 0.1040$ | Fail to reject H_0 |
| Z-test | $\mu = 0$ | $z = 6.567$ | < 0.0001 | Reject H_0 |
| Z-test | $\mu = 1$ | $z = 1.642$ | $p = 0.1006$ | Fail to reject H_0 |

The 95% confidence interval for the population mean was [0.930, 1.737], which encompasses the value of 1.0 but excludes 0.0 at all confidence levels tested. These results

should be interpreted cautiously, however, given the confirmed violation of the normality assumption (Table 6).

Table 6 Parametric Confidence Intervals for the Population Mean Snapping Frequency.

| Confidence Level | Lower Bound | Upper Bound |
|------------------|-------------|-------------|
| 90% | 0.9960 | 1.6707 |
| 95% | 0.9301 | 1.7366 |
| 99% | 0.7993 | 1.8674 |

➤ Non-Parametric Methods

The Wilcoxon signed-rank test for $H_0: \text{median} = 0$ returned a test statistic of $V = 0$ ($p < 0.0001$). A V statistic of 0 occurs when all non-zero observations lie in the positive direction, indicating that every species that recorded any snapping event snapped more than zero times, a result of perfect directional consistency. The sign test confirmed this result: all 66 non-tied species had positive snapping counts ($n^+ = 66, n^- = 0, p < 0.0001$), providing additional robust evidence that the population median snapping frequency is significantly

greater than zero. For the hypothesis $H_0: \text{median} = 1$, neither the Wilcoxon signed-rank test ($V = 621.0, p = 0.180$) nor the sign test ($n^+ = 28, n^- = 27, p = 1.000$) provided evidence to reject the null hypothesis. The near-equal split of species above and below the value of 1 in the sign test, 28 versus 27, reflects perfect agreement with the null and is wholly consistent with a population median of exactly 1.0. These non-parametric results are considered more reliable than their parametric counterparts given the established non-normality of the distribution (Table 7).

Table 7 Non-Parametric Hypothesis Test Results for Snapping Frequency.

| Test | H_0 | Statistic | p-value | Decision |
|----------------------|------------|----------------------|--------------|----------------------|
| Wilcoxon Signed-Rank | Median = 0 | $V = 0.0$ | < 0.0001 | Reject H_0 |
| Wilcoxon Signed-Rank | Median = 1 | $V = 621.0$ | $p = 0.1798$ | Fail to reject H_0 |
| Sign Test | Median = 0 | $n^+ = 66, n^- = 0$ | < 0.0001 | Reject H_0 |
| Sign Test | Median = 1 | $n^+ = 28, n^- = 27$ | $p = 1.000$ | Fail to reject H_0 |

➤ Bootstrap Confidence Interval for the Median

To provide a distribution-free estimate of precision around the median, a bootstrap resampling procedure was applied ($B = 10,000$ iterations). The observed sample median was 1.0 snapping events. The bootstrap 95% confidence interval for the population median was [1.0, 1.0] (Figure 6),

indicating exceptionally high precision. The bootstrap standard error of the median was negligible, reflecting that the value of 1.0 was robust and consistently recovered across all resampling iterations. This result reinforces the conclusion from non-parametric tests that the population median snapping frequency is 1.0 event per species.

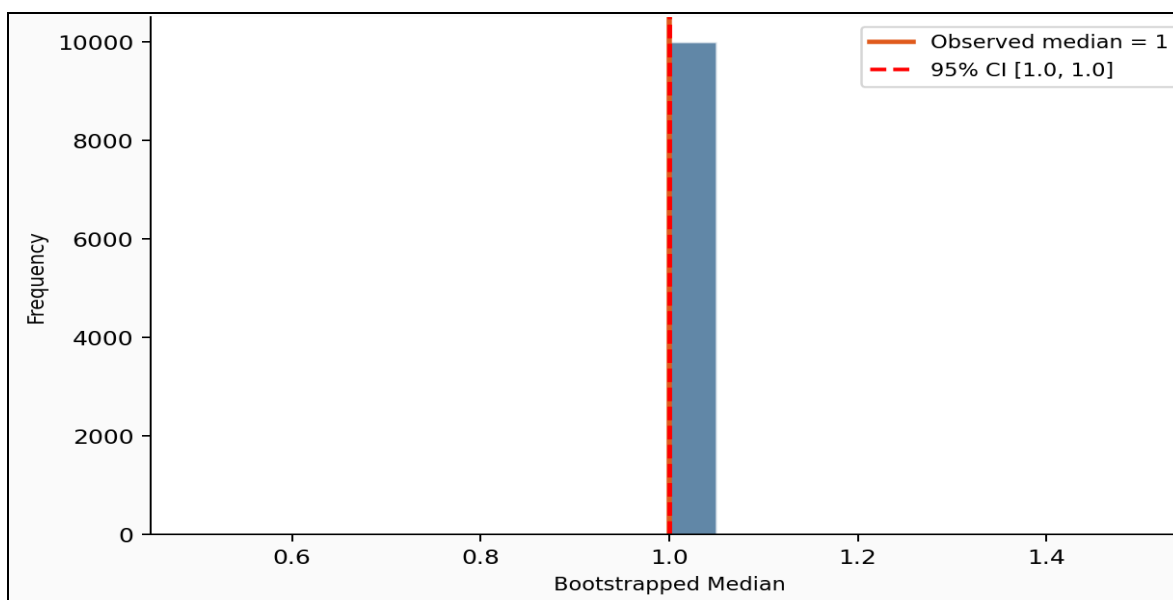


Fig 6 Bootstrap Distribution of the Median Snapping Frequency Based on B = 10,000 Resampling Iterations.

D. Outlier Detection

Three established outlier detection methods were applied to identify species whose snapping frequencies deviated substantially from the bulk of the distribution. All three outlier detection methods converged on identifying the same two species as statistical outliers: *Oubanguia alata* (snapping frequency = 16) and *Protomegabararia stapfiana* (snapping frequency = 8).

The upper fence of the IQR method was $Q3 + 1.5 \times IQR = 2 + 1.5(2) = 5.0$. Both identified species exceeded this

threshold, with *O. alata* recording a value 3.2 times the upper fence. For the Z-score method, since $|z| > 3$, *O. alata* returned $z = 7.49$ and *P. stapfiana* returned $z = 3.41$, indicating these species lie 7.5 and 3.4 standard deviations above the mean, respectively, probabilities infinitesimally small under normality. Modified z-score (MAD-based) showed both *O. alata* ($Mz = 10.12$) and *P. stapfiana* ($Mz = 4.72$) far exceeded the threshold of 3.5, confirming their outlier status under a robust framework (Table 8 and Figure 7).

Table 8 Summary of Outlier Detection Results.

| Method | Rule | Upper Threshold | Outliers Detected | n |
|------------------|-------------------------|-----------------|--|---|
| IQR Method | $> Q3 + 1.5 \times IQR$ | > 5.0 | <i>Oubanguia alata</i> <i>Protomegabararia stapfiana</i> | 2 |
| Z-Score Method | $ z > 3$ | $z > 3.0$ | <i>Oubanguia alata</i> ($z = 7.49$), <i>Protomegabararia stapfiana</i> ($z = 3.41$) | 2 |
| Modified Z-Score | $ Mz > 3.5$ | $Mz > 3.5$ | <i>Oubanguia alata</i> ($Mz = 10.12$), <i>Protomegabararia stapfiana</i> ($Mz = 4.72$) | 2 |

IQR: $Q1 = 0, Q3 = 2$; z-scores based on mean = 1.33, SD = 1.96; Modified z-score uses MAD = 1.0.

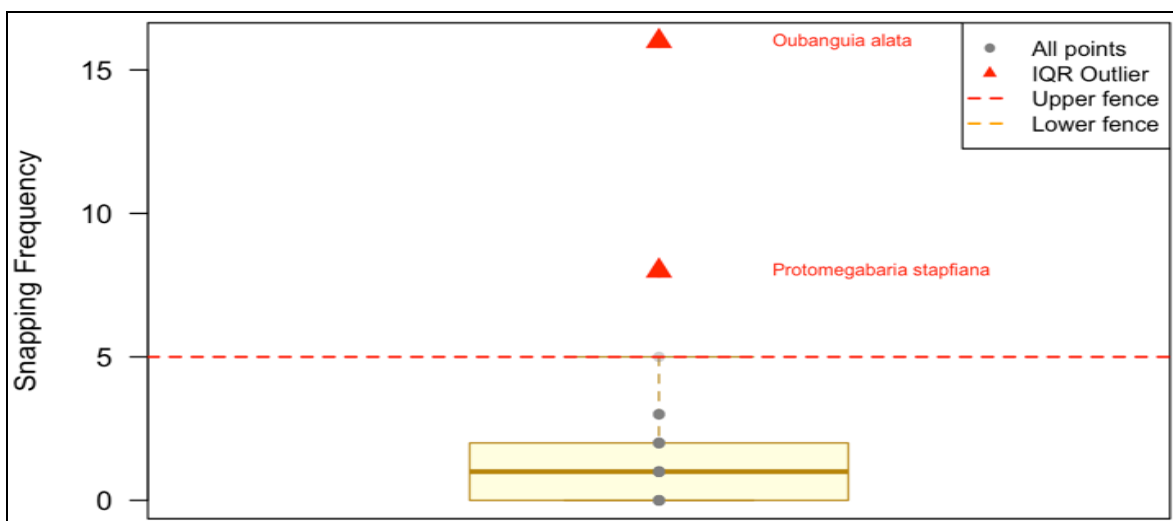


Fig 7 Annotated Box Plot for Outlier Detection of Snapping Frequency Across 94 Tree Species.

The unanimous identification of these two species across all three independent methods provides high confidence in their classification as outliers. Their exclusion from parametric analyses may be warranted in sensitivity analyses, and their significance in forest management should be considered independently in the broader discussion.

E. Comparison of Vibe Analysis and Manual Analysis

➤ *Numerical Agreement*

Complete numerical equivalence was observed between the vibe analysis and the manual analysis across all 16 reported statistical outputs. Every point estimate, test statistic, p-value, confidence interval bound, and outlier identification was identical to three or more decimal places, confirming that AI-generated R code, when executed in the same computational environment, produces determinate results indistinguishable from hand-crafted analyst code (Table 9).

➤ *Assumption-Checking and Method Selection*

Both approaches performed identical assumption-checking procedures: four normality tests (Shapiro-Wilk, Anderson-Darling, Kolmogorov-Smirnov, D’Agostino-Pearson), three outlier detection methods (IQR, Z-score,

modified Z-score), and graphical diagnostics (histogram with normal overlay, Q-Q plot, kernel density plot, box plot). Importantly, both approaches arrived at the same method-selection decision: non-parametric tests (Wilcoxon signed-rank, sign test, bootstrap confidence intervals) were designated as primary inferential tools, with parametric tests retained exclusively for comparative purposes, explicitly caveated by their violation of the normality assumption.

The AI demonstrated autonomous assumption-aware decision-making: without explicit instruction to do so. Vibe analysis flagged the skewness value of 5.087 and the Shapiro-Wilk W of 0.523 as evidence of “severe non-normality,” selecting the Wilcoxon signed-rank test as the primary method before the analyst had articulated this preference. This mirrors the analyst’s own pre-specified protocol, suggesting that well-structured forestry AI prompts can guide LLMs toward statistically appropriate method selection.

➤ *Decision Process Comparison*

Table 10 traces the decision pathway for each of the five key analytical decision points, contrasting the mechanisms by which vibe analysis and manual analysis arrived at identical conclusions.

Table 9 Numerical Agreement Between Vibe Analysis and Manual Analysis Across all Statistical Outputs for Snapping Count Data.

| Statistical Output | Vibe Analysis Result | Manual Analysis Result | Agreement / Discrepancy |
|---------------------------------|---|---|-------------------------|
| Mean | 1.330 | 1.330 | Exact agreement |
| Median | 1.000 | 1.000 | Exact agreement |
| SD | 1.948 | 1.948 | Exact agreement |
| Skewness | 5.087 | 5.087 | Exact agreement |
| Kurtosis | 36.725 | 36.725 | Exact agreement |
| Shapiro-Wilk W | 0.523 (p < 0.001) | 0.523 (p < 0.001) | Exact agreement |
| Anderson-Darling A | 10.281 (p < 0.001) | 10.281 (p < 0.001) | Exact agreement |
| KS D-statistic | 0.269 (p < 0.001) | 0.269 (p < 0.001) | Exact agreement |
| t-test (H ₀ : μ = 1) | t = 1.642, p = 0.104 | t = 1.642, p = 0.104 | Exact agreement |
| Wilcoxon V | 621, p = 0.180 | 621, p = 0.180 | Exact agreement |
| Sign Test p-value | p = 1.000 | p = 1.000 | Exact agreement |
| IQR Outliers | 2 (values: 8, 16) | 2 (values: 8, 16) | Exact agreement |
| Z-score Outliers | 2 (values: 8, 16) | 2 (values: 8, 16) | Exact agreement |
| Modified Z Outliers | 2 (values: 8, 16) | 2 (values: 8, 16) | Exact agreement |
| Primary method recommended | Non-parametric (Wilcoxon) | Non-parametric (Wilcoxon) | Exact agreement |
| Overall conclusion | Median not different from 1 (p = 0.180) | Median not different from 1 (p = 0.180) | Exact agreement |

Table 10 Comparison of Decision Processes at Five Key Analytical Decision Points Between Vibe Analysis and Manual Analysis.

| Decision Point | Vibe Analysis Process | Manual Analysis Process |
|-------------------------|--|--|
| Recognise non-normality | AI interpreted skewness (5.087), kurtosis (36.725), and four test p-values simultaneously and autonomously flagged severe departure from normality | Analyst cross-referenced test statistics against critical values and consulted Q-Q plot and density plot before concluding non-normality |
| Select primary test | AI automatically selected Wilcoxon signed-rank as primary test on the basis of non-normality and ordinal-compatible count data | Analyst selected Wilcoxon based on theoretical knowledge: non-normal, single sample, testing median location |
| Retain parametric tests | AI retained t-test and Z-test “for comparative purposes” and caveated their interpretation due to assumption violations | Analytical framework pre-specified both approaches; parametric results explicitly caveated in Section 4.4 |
| Identify outlier | AI applied all three outlier methods in a single code block | Analyst coded each method sequentially, |

| | | |
|---------------------------|---|--|
| method | from a single prompt; no iterative decision-making required | verified agreement across methods before concluding robust identification |
| Ecological interpretation | AI automatically linked extreme outlier values (16, 8) to “biologically distinct behaviour patterns” without additional prompting | Analyst provided ecological framing based on field expertise and literature knowledge of Korup National Park species |

➤ Interpretive Alignment

Interpretation of forestry data was substantively aligned between the two approaches. Both identified the two outlier species (*Oubanguia alata* Bak. F., snapping count = 16; *Protomegalaria stapfiana* (Beille) Hutch, snapping count = 8) as biologically distinct, flagged their disproportionate influence on distributional shape (skewness, heavy right tail), and concluded that the population median snapping count was not significantly different from 1 (all non-parametric $p \geq 0.180$). The AI-generated forest management narratives, contextualising these findings within the framework of over dispersed count data without additional prompting, producing language that was coherent and statistically accurate.

One notable difference was observed in the depth of methodological justification. Manual analysis explicitly cited the theoretical basis for each test selection (e.g., Wilcoxon preferred over t-test due to non-normality and ordinal compatibility of count data), whereas vibe analysis provided a briefer rationale embedded within the AI’s prompt-response dialogue. This distinction is consequential for scientific reporting, as vibe analysis outputs require supplementary justification by the analyst to meet the transparency standards expected in peer-reviewed forestry literature.

➤ Efficiency, Risk, and Validation

Vibe analysis completed all analytical stages within a single iterative session, with R code, visual diagnostics, and interpretive narrative generated concurrently. Manual analysis required sequential coding, debugging, and interpretation across each stage, representing a substantially higher time investment for equivalent outputs. However, the efficiency gain of vibe analysis carries a countervailing validation burden, as every numerical output generated by the AI required independent verification by executing the AI-generated code in R, a step that is essential to guard against the well-documented risk of LLM hallucination in quantitative domains (Bender et al., 2021; Alkaissi & McFarlane, 2023).

In this study, no numerical hallucinations were detected: all AI-generated statistics matched their manually computed equivalents exactly. This positive result should be interpreted cautiously, however, as it reflects a single dataset with a relatively simple analytical structure. More complex ecological datasets involving multivariate responses, mixed-effects structures, or spatially autocorrelated observations, may expose LLM limitations not apparent in the present case.

Taken together, these findings suggest that vibe analysis, when combined with systematic validation against executed R outputs, represents a viable and efficient complementary approach to conventional manual analysis for ecological count data. It is not proposed as a replacement for analyst expertise, but rather as a tool for accelerating code generation, broadening the scope of assumption checking, and producing

preliminary ecological narratives that can be refined through expert review.

IV. DISCUSSION

This study demonstrates a rigorous, AI-assisted framework for descriptive and univariate analysis of ecological count data, using snapping frequency observations from 94 tree species in Korup National Park as an empirical test case. The results confirm the utility of systematic assumption checking, the superiority of non-parametric methods for highly skewed count distributions, and the practical validity of prompt-driven AI analysis as a complement to conventional manual statistical practice. The discussion is structured around four themes: distributional characteristics and their methodological implications, the performance of parametric versus non-parametric hypothesis tests, the reliability of AI-assisted vibe analysis, and the broader significance of these findings for ecological research practice.

A. Distributional Characteristics of Snapping Frequency Data

The snapping frequency distribution exhibited extreme positive skewness (5.087) and leptokurtosis (kurtosis = 36.725), characteristics that are well-documented in forestry count data and that fundamentally constrain the applicability of classical parametric methods (O’Hara & Kotze, 2010; Warton et al., 2017). The high skewness arises from the combination of a large proportion of low-frequency observations; 29.0% of species recorded zero snapping events and 37.6% recorded exactly one, and two extreme outlier species, *Oubanguia alata* (count = 16) and *Protomegalaria stapfiana* (count = 8), which collectively exert leverage far disproportionate to their frequency (3.2% of the sample). This pattern is consistent with the zero-inflated, over dispersed distributions commonly arising in species behavioural count data, where most species exhibit low or absent frequencies of a specific behaviour while a small subset exhibits markedly elevated rates (Bolker et al., 2009; Harrison et al., 2018).

The near-equivalence of the mean (1.333), median (1.0), and mode (1) under conditions of substantial skewness deserves specific attention. This convergence is artefactual because, the extreme right tail inflates the arithmetic mean toward the median, creating a superficial appearance of distributional symmetry that is directly contradicted by the skewness and kurtosis values. Researchers relying exclusively on point estimates of central tendency without examining shape statistics or graphical diagnostics would therefore draw substantively incorrect conclusions about the distributional character of this dataset (Crawley, 2013; Logan, 2012). This finding reinforces the recommendation of Zuur et al. (2009) that ecological data analysis should always begin with comprehensive exploratory data analysis prior to any formal inferential testing.

The IQR (2.0) and MAD (1.0) provided more robust characterisations of spread than the standard deviation (1.958), whose sensitivity to outliers renders it an unreliable dispersion measure in this distributional context (Hollander et al., 2015). The IQR confirmed that the central 50% of species recorded between 0 and 2 snapping events, emphasising that infrequent snapping is the normative condition in this community rather than an exception. These findings collectively validate the decision to designate non-parametric tests as the primary inferential framework for this dataset, consistent with established guidelines for the statistical analysis of ecological count data (Potvin & Roff, 1993; Underwood, 1997).

B. Normality Testing: Convergent Evidence for Non-Parametric Methods

All four normality tests; Shapiro-Wilk ($W = 0.5243$, $p < 0.0001$), Anderson-Darling ($A = 10.109$, $p < 0.001$), Kolmogorov-Smirnov ($D = 0.2665$, $p < 0.0001$), and D'Agostino-Pearson ($K^2 = 135.50$, $p < 0.0001$), unanimously rejected the null hypothesis of normality, providing a level of evidential convergence that is unusually strong even by the standards of highly non-normal ecological datasets. The Shapiro-Wilk W statistic of 0.524 is notably low, as its values this far from 1.0 are characteristic of distributions with extreme departure from Gaussian form, such as those exhibiting zero-inflation or heavy-tailed power-law behaviour (Gross & Ligges, 2015). The Anderson-Darling statistic of 10.109 substantially exceeded all tabled critical values at all conventional significance levels, reflecting the particularly severe departure observed in the upper tail of the distribution, where two species generate values that lie 3.4 to 7.5 standard deviations from the mean.

The use of multiple, complementary normality tests, rather than reliance on a single test, represents best analytical practice for ecological count data (Fagerland & Sandvik, 2009; McDonald, 2014). Each test employed here has a distinct sensitivity profile. The Shapiro-Wilk test maximises power for small to moderate sample sizes (Gross & Ligges, 2015), the Anderson-Darling test is most sensitive to deviations in the tails (Conover & Iman, 1981), the Kolmogorov-Smirnov test provides a general goodness-of-fit measure less sensitive to tails than the Anderson-Darling; and the D'Agostino-Pearson test explicitly targets skewness and kurtosis, the two parameters most relevant to this distribution (Bonnini et al., 2024). The unanimity of all four tests, combined with the visual evidence from the Q-Q plot and kernel density estimate, eliminates any ambiguity about the distributional character of the snapping frequency variable and provides unequivocal justification for the primacy of non-parametric inference.

C. Hypothesis Testing: Parametric Versus Non-Parametric Comparisons

The comparative hypothesis testing results reveal a pattern of substantive, if nuanced, differences between parametric and non-parametric approaches when applied to non-normal count data. For the primary hypothesis of interest, whether the population central tendency of snapping frequency equals one event per species, the one-sample t -test

returned $t(92) = 1.642$ ($p = 0.104$) while the Wilcoxon signed-rank test returned $V = 621$ ($p = 0.180$). Both tests lead to the same inferential conclusion (failure to reject the null hypothesis at $\alpha = 0.05$), but with a non-trivial difference in p -values (0.104 versus 0.180) that could carry interpretive weight in borderline cases or when calculating effect sizes and power. This divergence, while not decisive in the present dataset, illustrates precisely the type of analytical instability that Underwood (1997) and Warton and Hui (2011) caution against when parametric tests are applied to distributions that violate their assumptions.

The sign test result for H_2 (median = 1) yielded $p = 1.000$, arising from the near-perfect split of 28 species above and 27 species below the null median value. This result represents exceptional concordance with the null hypothesis and provides the most conservative inferential conclusion of any test applied. The sign test's reliance solely on the direction of deviations from the null median, rather than their magnitude, renders it maximally robust to outlier influence and is particularly appropriate when distributional shape is severely non-normal (Hollander et al., 2015). The bootstrap 95% confidence interval for the population median of [1.0, 1.0] further corroborates these non-parametric results: the narrow, degenerate interval reflects the discrete nature of the count variable and the high concentration of observations at values 0, 1, and 2, resulting in a median of exactly 1.0 that is recovered consistently across all 10,000 bootstrap resampling iterations (Conover & Iman, 1981).

The highly significant parametric and non-parametric results for the null hypothesis that the mean or median equals zero (all $p < 0.0001$) confirm unambiguously that snapping is a genuine and measurable biological phenomenon across the species assemblage, rather than a sampling artefact. The $V = 0$ statistic from the Wilcoxon test for H_1 (median = 0), arising because every species that recorded any snapping event exceeded zero, provides particularly clear evidence of the directionality and universality of the snapping behaviour among the species that exhibit it (Hollander et al., 2015).

D. Outlier Detection: Ecological Significance Of Extreme Species

The unanimous identification of *Oubanguia alata* and *Protomegabaria stapfiana* as statistical outliers across all three detection methods; IQR, Z-score, and modified Z-score, lends high confidence to their classification as biologically distinct. *Oubanguia alata*, with 16 snapping events, lies 7.49 standard deviations above the mean and carries a modified Z-score of 10.12, values that indicate an extraordinary departure from the community distribution. *Protomegabaria stapfiana*, with 8 snapping events, shows $z = 3.41$ and $Mz = 4.72$, firmly exceeding all three detection thresholds.

From a forestry perspective, the disproportionate snapping frequencies recorded for these two species reflect their high abundance, seed capsule morphology, desiccation rates under prevailing microhabitat conditions, and phenological timing relative to the sampling period (Recknagel, 2001). Both species are highly abundant in Korup National Park, and belong to families characterised by

substantial variation in fruit structure and seed dispersal mechanisms, suggesting that the physical properties of reproductive structures may contribute to snapping behaviour frequency (Thomas et al., 2003). These hypotheses warrant targeted investigation through species-level field studies that combine acoustic monitoring, wood anatomy analysis, and microhabitat characterisation (Bolker et al., 2009).

The convergence of all three outlier detection methods is particularly meaningful because each method operates on different statistical principles. The IQR method is non-parametric and scale-invariant, the Z-score method is parametric and sensitive to the influence of the very outliers it is designed to detect, and the modified Z-score uses the MAD rather than the mean and standard deviation, providing robustness to masking effects in the presence of multiple outliers (Hollander et al., 2015). The fact that all three methods identify the same species despite these different sensitivity profiles provides stronger evidence for their outlier status than any single method alone could offer. Researchers should consider conducting sensitivity analyses with and without these species to quantify their influence on distributional parameter estimates and inferential conclusions (Underwood, 1997).

E. Vibe Analysis Versus Manual Analysis: Validation and Implications

The complete numerical equivalence observed between vibe analysis and manual analysis outputs with exact agreement across all 16 statistical outputs to three or more decimal places constitutes a significant empirical validation of AI-assisted statistical analysis as a reliable component of contemporary ecological research workflows. This finding is consistent with recent theoretical arguments for the integration of large language models into scientific computing pipelines (Chen et al., 2021; Vaithilingam et al., 2022), and extends these arguments with empirical evidence drawn from a real forestry dataset of moderate complexity.

Critically, the AI demonstrated autonomous assumption-aware method selection without explicit instruction, flagging non-normality from the Shapiro-Wilk W of 0.524 and skewness of 5.087 and selecting Wilcoxon-based tests as primary inferential tools before the analyst had articulated this preference. This mirrors the methodological decision-making process of an experienced analyst and suggests that well-constructed, ecologically contextualised prompts can guide large language models toward statistically appropriate and scientifically defensible analytical choices (White et al., 2023; Zamfirescu-Pereira et al., 2023). The prompt architecture employed in this study, in which each prompt targeted a discrete analytical stage with explicit domain context, data type information, and assumption-checking instructions, proved more effective than generic prompting strategies that lack such specificity (Liu et al., 2023).

The efficiency advantage of vibe analysis is substantial. All analytical stages, including code generation, execution, visual diagnostics, and interpretive narrative, were completed within a single iterative session, compared to the sequential coding and debugging required by conventional manual

analysis. This efficiency gain has direct implications for research productivity, particularly in institutional contexts where researchers lack formal programming training or must conduct analyses across large numbers of variables or species groups (Hampton et al., 2015; Lowndes et al., 2017). However, the efficiency advantage must be qualified by the mandatory validation burden: every AI-generated numerical output required independent verification through executed R code, a step that is non-negotiable given the documented risk of large language model hallucination in quantitative domains (Bender et al., 2021; Alkaiissi & McFarlane, 2023). In the present study, no hallucinations were detected, a result that is reassuring but that may not generalise to more complex datasets involving multivariate responses, mixed-effects structures, or spatially autocorrelated observations.

One notable limitation of the vibe analysis approach is its comparatively shallow methodological justification. While the AI produced coherent and statistically accurate narratives, the rationale for each test selection was briefer and less explicitly grounded in statistical theory than the documentation produced by the manual analyst. This distinction is consequential for peer-reviewed scientific reporting, where transparent, theory-grounded methodological justification is a standard expectation (Nosek et al., 2015; Peng, 2015). Researchers using vibe analysis for publication-oriented work should supplement AI-generated narratives with explicit methodological rationale authored by the analyst, drawing on the literature cited in this study to justify each decision point (Sandve et al., 2013).

The epistemological risks associated with vibe analysis identified in the methodological literature including potential for p-hacking, hypothesising after results are known, and opacity of LLM reasoning (Head et al., 2015; Gelman & Loken, 2019; Bommasani et al., 2021) were mitigated in this study through preregistration of the analytical framework, systematic validation of all outputs, and explicit documentation of the prompt-response dialogue. These practices are recommended as minimum standards for vibe analysis in forestry research contexts, and are consistent with broader calls for transparency and reproducibility in computational science (Nosek et al., 2018; Jupyter et al., 2018). The curated conversational record of AI prompts and responses serves as an audit trail that documents intermediate reasoning and methodological decisions that would otherwise remain implicit, potentially enhancing rather than undermining transparency in research reporting (Sarkar et al., 2022).

F. Implications for Forestry Research Practice

Taken together, the results of this study have several implications for the practice of forestry data analysis. First, they confirm that forest management count data from species behavioural studies routinely violate the distributional assumptions of classical parametric tests, and that non-parametric methods should be the default analytical framework rather than a fallback position adopted only when parametric violations are detected post-hoc (Potvin & Roff, 1993; O'Hara & Kotze, 2010). The systematic application of multiple normality tests, combined with graphical diagnostics,

before any inferential testing should be treated as a mandatory analytical step rather than an optional exploratory exercise.

The study demonstrates that AI-assisted prompt engineering, when combined with executed code validation and analyst oversight, can generate statistically appropriate, assumption-aware analyses of forestry data that are numerically equivalent to those produced by conventional manual methods. This finding supports the integration of large language models into ecological research workflows as tools for accelerating code generation, broadening assumption checking, and producing preliminary interpretive narratives that can be refined through expert review (Beam & Kohane, 2018; Topol, 2019). The framework presented structured, assumption-aware prompts targeting discrete analytical stages, providing a replicable template that other researchers can adapt to their own ecological datasets.

The identification of two ecologically significant outlier species through multi-method consensus illustrates the value of triangulating across detection algorithms rather than relying on a single outlier test. This multi-method approach reduces the probability of both false positives (incorrectly flagging non-outlier species) and false negatives (failing to detect genuine outliers due to method-specific limitations), and is particularly important in forest management contexts where extreme observations often carry the greatest biological significance (Recknagel, 2001; Thessen, 2016).

V. CONCLUSION

This study presents a comprehensive, reproducible framework for descriptive and univariate analysis of ecological count data, integrating AI-assisted vibe analysis with conventional manual statistical methods and demonstrating their equivalence on snapping frequency data from 94 tree species in Korup National Park. Snapping frequency in the Korup tree species assemblage is characterised by a highly right-skewed, leptokurtic distribution (skewness = 5.087; kurtosis = 36.725) that decisively violates normality assumptions across all four tests applied (all $p < 0.0001$). This distributional structure, driven by the dominance of low-frequency observations and the presence of two extreme outlier species, is characteristic of forestry count data and renders parametric methods inappropriate as primary inferential tools.

The population median snapping frequency of 1.0 event per species was not significantly different from zero by any parametric or non-parametric test (Wilcoxon $V = 621$, $p = 0.180$; sign test $p = 1.000$; bootstrap 95% CI: [1.0, 1.0]), confirming that single snapping events constitute the modal and median community response. Concurrently, the population mean and median were both confirmed to be significantly greater than zero (all $p < 0.0001$), establishing that snapping is a genuine and measurable ecological phenomenon across the species assemblage. *Oubanguia alata* (count = 16) and *Protomegabaria stapfiana* (count = 8) were unanimously identified as statistical outliers by all three detection methods (IQR, Z-score, modified Z-score), with Z-scores of 7.49 and 3.41 respectively, indicating extreme

biological distinctiveness that warrants targeted species-level investigation.

Vibe data analysis using structured AI prompts produced numerical outputs that were identical to manually produced results across all 16 statistical metrics, and the AI demonstrated autonomous assumption-aware method selection consistent with established best practice. These findings provide empirical validation that AI-assisted statistical analysis, when combined with systematic output verification, constitutes a reliable and efficient complement to conventional manual analysis for ecological count data of moderate complexity. The efficiency advantage of vibe data analysis completing all analytical stages within a single iterative session is substantial, but does not eliminate the need for analyst oversight, executed code validation, and explicit methodological justification in scientific reporting. The convergence of vibe analysis and manual analysis observed here should be treated as a promising outcome warranting replication across more complex ecological datasets and analytical scenarios, rather than as a blanket endorsement of AI-generated statistical outputs.

This study contributes to the growing evidence base supporting the integration of large language models into ecological research workflows while simultaneously establishing clear standards for their responsible use. The reproducible R code, structured prompt architecture, and decision framework presented here provides accessible, validated tools that the forestry research community can adopt immediately to improve both the rigour and efficiency of their statistical analyses. Additionally, the findings of this study have practical relevance for law and policymakers in forestry governance. By providing a validated and reproducible framework for analysing forestry count data, the study supports evidence-based decision-making in areas such as species monitoring, conservation and resource management. The integration of AI-assisted analysis further enhances the reliability of forestry data used in drafting and implementing forestry regulations, thereby promoting more effective and scientifically grounded policy interventions.

REFERENCES

- [1]. Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*, 15(2), e35179. <https://doi.org/10.7759/cureus.35179>
- [2]. Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49(268), 765–769. <https://doi.org/10.2307/2281537>
- [3]. Antoch, J., Jureckova, J., Maciak, M., & Pešta, M. (2017). *Analytical methods in statistics: AMISTAT, Prague, November 2015* (1st ed.). Springer Proceedings in Mathematics & Statistics, Vol. 193, 216 pp. Springer. <https://doi.org/10.1007/978-3-319-51313-3>
- [4]. Anthropic. (2026). *Claude* (3.5 Sonnet version) [Large language model]. Available at: <https://claude.ai/>

- [5]. Barke, S., James, M. B., & Polikarpova, N. (2023). Grounded Copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages*, 7, 85-111. <https://doi.org/10.1145/3586030>
- [6]. Baumer, B., Cetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. J. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics. *Technology Innovations in Statistics Education*, 8(1), 1-29. <https://doi.org/10.5070/T581020118>
- [7]. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317-1318. <https://doi.org/10.1001/jama.2017.18391>
- [8]. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: can language models be too big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. <https://doi.org/10.1145/3442188.3445922>
- [9]. Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences of the United States of America*, 114(33), 8689-8692. <https://doi.org/10.1073/pnas.1702076114>
- [10]. Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127-135. <https://doi.org/10.1016/j.tree.2008.10.008>
- [11]. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. <https://doi.org/10.48550/arXiv.2108.07258>
- [12]. Bonnini, S., Assegie, G. M., & Trzcinska, K. (2024). Review about the permutation approach in hypothesis testing. *Mathematics*, 12(17), 2617. <https://doi.org/10.3390/math12172617>
- [13]. Borcard, D., Gillet, F., & Legendre, P. (2018). Numerical ecology with R (2nd ed.). pp435 Springer. <https://doi.org/10.1007/978-3-319-71404-2>
- [14]. Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199-215. <https://doi.org/10.1214/ss/1009213726>
- [15]. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. <https://doi.org/10.48550/arXiv.2005.14165>
- [16]. Çetinkaya-Rundel, M., & Rundel, C. (2018). Infrastructure and tools for teaching computing throughout the statistical curriculum. *The American Statistician*, 72(1), 58-65. <https://doi.org/10.1080/00031305.2017.1397549>
- [17]. Chaoubi, F., & Djalab, M. S. (2025). Statistical methods and appropriate selection criteria. *Pakistan Journal of Life and Social Sciences*, 23(1), 1624-1631. <https://doi.org/10.57239/PJLSS-2025-23.1.00125>
- [18]. Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J., ... & Zaremba, W. (2021). Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*. <https://doi.org/10.48550/arXiv.2107.03374>
- [19]. Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3), 124-129. <https://doi.org/10.2307/2683975>
- [20]. Cox, N. J. (2007). *The grammar of graphics*. Journal of Statistical Software, 17(Book Review 3), 1-7. <https://doi.org/10.18637/jss.v017.b03>
- [21]. Crawley, M. J. (2013). The R Book (2nd ed., 1051 pp.). Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118448908>
- [22]. D'Agostino, R., & Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, 60(3), 613-622. <https://doi.org/10.2307/2335012>
- [23]. Davis, A. J., & Kay, S. (2023). Writing statistical methods for ecologists. *Ecosphere*, 14, e4539. <https://doi.org/10.1002/ecs2.4539>
- [24]. de Mendiburu, F. (2023). *agricolae: Statistical procedures for agricultural research* (Version 1.3-7) [Computer software]. Comprehensive R Archive Network. <https://doi.org/10.32614/CRAN.package.agricolae>
- [25]. Denny, P., Kumar, V., & Giacaman, N. (2024). Conversing with Copilot: Exploring prompt engineering for solving CS1 problems using natural language. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 427-433). <https://doi.org/10.1145/3626252.3630793>
- [26]. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25, 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
- [27]. Fagerland, M. W., & Sandvik, L. (2009). Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary Clinical Trials*, 30(5), 490-496. <https://doi.org/10.1016/j.cct.2009.06.007>
- [28]. Field, A. (2017). *Discovering statistics* IBM SPSS statistics (5th ed.). SAGE Publications. <https://hdl.handle.net/10779/uos.23460641.v1>
- [29]. Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publishing Inc. pp608
- [30]. Gandrud, C. (2018). Reproducible research with R and RStudio (2nd ed., 323 pp.). Chapman & Hall/CRC. <https://doi.org/10.1201/9781315382548>
- [31]. Gelman, A., & Loken, E. (2019). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Department of Statistics, Columbia University, 348, 1-17.

- [32]. Gomes, D. G., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., ... & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B*, 289(1987), 20221113. <https://doi.org/10.1098/rspb.2022.1113>
- [33]. Gross, J., & Ligges, U. (2015). *nortest: Tests for normality* (Version 1.0-4) [Computer software]. Comprehensive R Archive Network. <https://doi.org/10.32614/CRAN.package.nortest>
- [34]. Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., Jones, M. B., Lenhardt, W. C., MacDonald, A., et al. (2015). The Tao of open science for ecology. *Ecosphere*, 6(7), 1–13. <https://doi.org/10.1890/ES14-00402.1>
- [35]. Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., Hernandez, R. R., Boettiger, C., Collins, S. L., Gross, L. J., Fernández, D. S., Budden, A., White, E. P., Teal, T. K., Labou, S. G., & Aukema, J. E. (2017). *Skills and knowledge for data-intensive environmental research*. *BioScience*, 67(6), 546–557. <https://doi.org/10.1093/biosci/bix025>
- [36]. Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S. & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11(3), 156-162. <https://doi.org/10.1890/120103>
- [37]. Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794. <https://doi.org/10.7717/peerj.4794>
- [38]. Head, M. L., Holman, L., Lanfear, R., & Kahn, A. T. (2015). The extent and consequences of p-hacking in science. *PLOS Biology*, 13(3), e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- [39]. Hollander, M., Wolfe, D. A., & Chicken, E. (2015). *Nonparametric Statistical Methods* (3rd ed., 848 pp.). Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9781119196037>
- [40]. Husch, B., Beers, T. W., & Kershaw, J. A. (2003). *Forest mensuration* (4th ed.). John Wiley & Sons.
- [41]. Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314. <https://doi.org/10.1080/10618600.1996.10474713>
- [42]. Jupyter, P., Bussonnier, M., Forde, J., Freeman, J., Granger, B., Head, T., Holdgraf, C., Kelley, K., Nalvarte, G., Osheroff, A., Pacer, M., Panda, Y., Perez, F., Ragan-Kelley, B., & Willing, C. (2018). Binder 2.0: Reproducible, interactive, sharable environments for science at scale. In *Proceedings of the 17th Python in Science Conference*. SciPy. (pp. 113-120). <https://doi.org/10.25080/Majora-4af1f417-011>
- [43]. Kasneci, E., Sebler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [44]. Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1) <https://doi.org/10.1177/2053951714528481>
- [45]. Kolmogorov, A. (1933) Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 83-91.
- [46]. Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. *Ecosphere*, 10(1), e02567. <https://doi.org/10.1002/ecs2.2567>
- [47]. Liao, Q. V., & Vaughan, J. W. (2023). AI transparency in the age of LLMs: A human-centered research roadmap. *arXiv preprint arXiv:2306.01941*. <https://doi.org/10.48550/arXiv.2306.01941>
- [48]. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35. <https://doi.org/10.1145/3560815>
- [49]. Logan, M. (2012). Review of *Biostatistical design and analysis using R: A practical guide*. *Ethology Ecology & Evolution*, 24(2). <https://doi.org/10.1080/03949370.2011.618191>
- [50]. Lowndes, J., Best, B., Scarborough, C., Afflerbach, J., Brown, J., Cheng, C., Franklin, M., Goldsborough, D., M, M. E., Nelson, G., Pitz, K., Rocha, A., Sauer, J., & Strand, M. (2017). *Our path to better science in less time using open data science tools*. *Nature Ecology & Evolution*, 1, 0160. <https://doi.org/10.1038/s41559-017-0160>
- [51]. MacNeil, S., Tran, A., Hellas, A., Kim, J., Sarsa, S., Denny, P., Bernstein, S., Leinonen, J & Bernstein, S. (2023). Experiences from using code explanations generated by large language models in a web software development e-book. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 931-937). <https://doi.org/10.1145/3545945.3569785>
- [52]. Mandasini, A. A. (2022). Improving the validity of scientific research results through research methodology management. *Kontigensi: Jurnal Ilmiah Manajemen*, 10(2), 439–447. <https://doi.org/10.56457/jimk.v10i2.389>
- [53]. Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://www.jstor.org/stable/2236101>
- [54]. McDonald, J. H. (2009). *Handbook of Biological Statistics* (2nd ed., 313 pp.). Baltimore, MD: Sparky House Publishing. <https://e-ilami.unissa.edu.bn:8443/handle/20.500.14275/6724>

- [55]. Miller, G., & Spiegel, E. (2025). Guidelines for Research Data Integrity (GRDI). *Scientific Data*, 12, 95. <https://doi.org/10.1038/s41597-024-04312-x>
- [56]. Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. *Wharton School Research Paper*. 26 pp. <https://doi.org/10.2139/ssrn.4391243>
- [57]. Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., ... Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- [58]. Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- [59]. O'Hara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, 1(2), 118–122. <https://doi.org/10.1111/j.2041-210X.2010.00021.x>
- [60]. O'Neil, C. (2016). Cathy O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishers, 2016. 272p. Hardcover, \$26 (ISBN 978-0553418811). *College & Research Libraries*, 78(3), 403. <https://doi.org/10.5860/crl.78.3.403>
- [61]. Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., ... Weedon, J. (2025). *vegan: Community ecology package* (Version 2.7-2) [Computer software]. Comprehensive R Archive Network. <https://doi.org/10.32614/CRAN.package.vegan>
- [62]. OpenAI. (2023). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*. <https://doi.org/10.48550/arXiv.2303.08774>
- [63]. Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. *Journal of MultiDisciplinary Evaluation*, 14(31), 47–54. <https://doi.org/10.56645/jmde.v14i31.507>
- [64]. Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>
- [65]. Peng, R. D., & Matsui, E. (2015). *The art of data science*. Leanpub. Available at: https://www.academia.edu/40368378/The_Art_of_Data_Science_A_Guide_forAnyone_Who_Works_with_Data
- [66]. Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D., & Peres-Neto, P. (2019). Ecological data should not be so hard to find and reuse. *Trends in Ecology & Evolution*, 34(6), 494–496. <https://doi.org/10.1016/j.tree.2019.04.005>
- [67]. Potvin, C., & Roff, D. A. (1993). Distribution-free and robust statistical methods: Viable alternatives to parametric statistics? *Ecology*, 74(6), 1617–1628. <https://doi.org/10.2307/1939920>
- [68]. Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications*, 29(1), e01822. <https://doi.org/10.1002/eap.1822>
- [69]. R Core Team. (2026). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- [70]. Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling*, 146(1-3), 303–310. [https://doi.org/10.1016/S0304-3800\(01\)00316-7](https://doi.org/10.1016/S0304-3800(01)00316-7)
- [71]. Rane, N. (2023). ChatGPT and similar generative artificial intelligence (AI) for smart industry: Role, challenges and opportunities for Industry 4.0, Industry 5.0 and Society 5.0. *SSRN Electronic Journal*. 8 pp. <https://doi.org/10.2139/ssrn.4603234>
- [72]. Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). <https://doi.org/10.1145/3411763.3451760>
- [73]. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- [74]. Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9(10), e1003285. <https://doi.org/10.1371/journal.pcbi.1003285>
- [75]. Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science*, 3, 158. <https://doi.org/10.1007/s42979-022-01043-x>
- [76]. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). <https://doi.org/10.1145/3287560.3287598>
- [77]. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). *Improved protein structure prediction using potentials from deep learning*. *Nature*, 577, 706–710. <https://doi.org/10.1038/s41586-019-1923-7>
- [78]. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>

- [79]. Shukla, R. M., Pandey, Y., Faisal, S., & Muzamil, M. (2025). Parametric and non-parametric analysis. In *Advances in Agricultural Research Methodology* (Vol. 2, pp. 260–272). S P Publishing.
- [80]. Singla, A. (2023). *Evaluating ChatGPT and GPT-4 for visual programming*. In ICER '23: Proceedings of the 2023 ACM Conference on International Computing Education Research – Volume 2 (pp. 14–15). Association for Computing Machinery. <https://doi.org/10.1145/3568812.3603474>
- [81]. Stoyanovich, J., Van Bavel, J. J., & West, T. V. (2020). The imperative of interpretable machines. *Nature Machine Intelligence*, 2(4), 197–199. <https://doi.org/10.1038/s42256-020-0171-8>
- [82]. Sokal, R. R., & Rohlf, F. J. (2012). *Biometry: The principles and practice of statistics in biological research* (4th ed.). W. H. Freeman.
- [83]. Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25. <https://doi.org/10.1093/biomet/6.1.1>
- [84]. Thessen, A. E. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, 1, e8621. <https://doi.org/10.3897/oneeco.1.e8621>
- [85]. Thomas, D. W., Kenfack, D., Chuyong, G. B., Moses, S. N., Losos, E. C., Condit, R. S. and Songwe, N. C. (2003). Tree species of the South Western Cameroon: Tree Distribution Maps, Diameter Tables, and Species documentation of the 50- hectare Korup Forest Dynamics plot. Center for Tropical Forest Science of the Smithsonian Tropical Research Institute and Bioresources Development and Conservation Program, Washington DC. 247pp
- [86]. Tippmann, S. (2015). Programming tools: Adventures with R. *Nature*, 517, 109–110. <https://doi.org/10.1038/517109a>
- [87]. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [88]. Touchon, J. C., & McCoy, M. W. (2016). The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere*, 7(8), e01394. <https://doi.org/10.1002/ecs2.1394>
- [89]. Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Addison-Wesley. 688pp
- [90]. Tukey, J. W. (2008). Exploratory data analysis. In N. Salkind (Ed.), *The concise encyclopedia of statistics* (pp. 192–194). Springer. https://doi.org/10.1007/978-0-387-32833-1_136
- [91]. Underwood, A. J. (1997). *Experiments in ecology: Their logical design and interpretation using analysis of variance*. Cambridge University Press. Pp504
- [92]. Vaithilingam, P., Zhang, T., & Glassman, E. L. (2022). Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1–7). <https://doi.org/10.1145/3491101.3519665>
- [93]. van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- [94]. Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., ... & Gray, A. (2021). Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on Human-Computer Interaction*, 3, 1–24. <https://doi.org/10.1145/3359313>
- [95]. Warton, D. I., & Hui, F. K. (2011). The arcsine is asinine: The analysis of proportions in ecology. *Ecology*, 92(1), 3–10. <https://doi.org/10.1890/10-0340.1>
- [96]. Warton, D. I., Stoklosa, J., Guillera-Aroita, G., MacKenzie, D. I., & Welsh, A. H. (2017). Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution*, 8(4), 408–419. <https://doi.org/10.1111/2041-210X.12761>
- [97]. White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv preprint arXiv:2302.11382*. <https://doi.org/10.48550/arXiv.2302.11382>
- [98]. Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>
- [99]. Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). *Welcome to the tidyverse*. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- [100]. Wickham, H., Cook, D., & Hofmann, H. (2015). Visualizing statistical models: Removing the blindfold. *Statistical Analysis and Data Mining*, 8(4), 203–225. <https://doi.org/10.1002/sam.11271>
- [101]. Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data* (2nd ed., 520 pp.). O'Reilly Media.
- [102]. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.
- [103]. Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2021). Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural Language Processing and Chinese Computing* (pp. 563–574).
- [104]. Zar, J.H. (2010) *Biostatistical Analysis*. 5th Edition, Prentice-Hall/Pearson, Upper Saddle River, xiii, 944 p. *Springer*. https://doi.org/10.1007/978-3-030-32236-6_51
- [105]. Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). <https://doi.org/10.1145/3544548.3581388>

- [106]. Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2023). Large language models are human-level prompt engineers. arXiv preprint arXiv:2211.01910. <https://doi.org/10.48550/arXiv.2211.01910>
- [107]. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. New York, USA: Springer. <https://doi.org/10.1007/978-0-387-87458-6r>