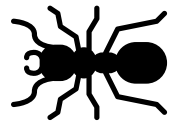
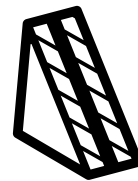


The Sandwich Ant Experiment

Statistical Insights into
Food Preference in Ants



Author: Seyed Ahmad Ahmadi

Date: April 21, 2025

TABLE OF CONTENTS

1. Introduction

2. Detailed Description of the Problem

3. Methods

a. Descriptive Statistics

b. Kruskal-Wallis H-test

c. Mann-Whitney U Test

4. Evaluation

a. Descriptive Analysis

b. Statistical Hypothesis Test

5. Summary

6. Bibliography

7. Appendix

INTRODUCTION

Ants are the most common insects in both outdoor and indoor environments, where they are often considered pests. Their primary goal is finding food. Understanding what attracts ants to food items can inform pest control strategies, food storage practices, and food preparation choices in ant-prone areas, such as picnic areas, kitchens, and outdoor dining spaces. This study focuses on sandwiches, a popular food item, and investigates how their components—specifically bread type, topping, and the presence of butter—affect ant attraction. The motivation for this research stems from the need to reduce food waste and prevent ant infestations, which can pose health risks by spreading pathogens (Hölldobler & Wilson, 1990).

The primary research questions are:

1. Does the type of bread significantly influence the number of ants attracted?
2. Does the topping significantly influence the number of ants attracted?
3. Does the use of butter significantly influence the number of ants attracted?

To answer these questions, we use a dataset containing ant counts for various sandwich combinations. Our approach involves descriptive statistics to explore the data, followed by nonparametric statistical tests (Kruskal-Wallis H-test and Mann-Whitney U test) to assess significance. The main result indicates that topping type significantly influences ant attraction, with Ham and gherkins being the most attractive, especially when butter is used. Bread type and butter show no significant effect, though butter may increase ant attraction slightly.

This report is structured as follows: Section 2 describes the problem and dataset, Section 3 outlines the statistical methods, Section 4 presents the evaluation, Section 5 summarizes the findings, and Sections 6 and 7 provide the bibliography and appendix, respectively.

DETAILED DESCRIPTION OF THE PROBLEM

The task is to determine whether bread type, topping, and butter significantly affect the number of ants attracted to sandwiches and to identify which bread type and topping attract the most ants. The dataset, stored in `sandwich.csv`, contains 48 observations with the following variables:

- `antCount`: Number of ants which are attracted to sandwiches (numeric, continuous).
- `bread`: Type of bread (categorical, nominal; values: Whole Grain, White, Rye, Multi Grain).
- `topping`: Type of topping (categorical, nominal; values: Ham and gherkins, Peanut butter, Yeast spread).
- `butter`: is butter used in sandwiches or not (categorical, nominal; values: yes, no).

The data appears to be collected from a experiment where sandwiches with different combinations of bread, topping, and butter were used for the experiment , and the number of ants attracted was recorded. The dataset contains no missing values, as confirmed by an initial check using Python's pandas library (`data.isna().sum()` returned 0 for all columns). However, the sample sizes vary across groups—for For example, there are 12 observations for Rye bread but only 9 for ham and gherkins with butter. While this imbalance is common in observational data, it could impact the statistical power of the analysis.

METHODS

This section describes the statistical methods used, including mathematical definitions, explanations, and literature references.

Descriptive Statistics

We summarized the dataset using descriptive statistics, focusing on the median and interquartile range (IQR) because ant count data may not follow a normal distribution.

Median: The middle value of an ordered dataset. If the sample size (n) is odd, the median is the central value; if even, it's the average of the two middle values. For example:

- Odd n : Median = middle observation.
- Even n : Median = (observation at $n/2$ + observation at $(n/2 + 1)$) / 2.

This estimator provides the 50th percentile of the empirical distribution while minimizing sensitivity to extreme values.

(Conover, 1999, p. 21).

Interquartile Range (IQR): Measures spread by calculating the difference between the 75th percentile (Q3) and the 25th percentile (Q1). Unlike standard deviation, the IQR is less affected by extreme values, making it reliable for skewed data.

$$IQR = Q3 - Q1$$

Key properties:

- Captures the middle 50% of the data distribution
- Resistant to outliers (breakdown point = 25%)
- Serves as the basis for outlier detection (Tukey's fences: $Q1 - 1.5 \times IQR$ to $Q1 + 1.5 \times IQR$, $Q3 - 1.5 \times IQR$ to $Q3 + 1.5 \times IQR$)

(Hollander et al., 2013).

Both estimators are particularly appropriate for count data, which often exhibit:

- Right-skewness (Poisson-like distributions)
- Zero-inflation
- Overdispersion relative to parametric assumptions

Kruskal-Wallis H-test

The Kruskal-Wallis H-test serves as a distribution-free method for comparing three or more independent groups. This test evaluates whether the groups originate from populations with identical medians, making it particularly valuable when parametric assumptions are violated.

- **Methodology Overview:**

- a. **Data Preparation:**

- Combine all observations across groups
 - Assign ranks to each value (1 for smallest)
 - Apply average ranks for tied values

- b. **Test Statistic Calculation:**

- The H-statistic compares the observed rank sums against expected values under the null hypothesis:

$$H = \left[\frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right] - 3(N+1)$$

Where : N = totalsamplesize - k = numberofgroups - R_i = sumofranksforgroup i - n_i = samplesizeofgroup i

- c. **Statistical Inference:**

- For adequate sample sizes (typically $n \geq 5$ per group), H follows a χ^2 distribution with $k-1$ degrees of freedom
 - Significant results indicate at least one group differs from the others

(Conover, 1999, p. 288).

- **Advantages:** Does not assume normality; suitable for ordinal or non-normal continuous data.
- **Disadvantages:** Less powerful than ANOVA if data is normal; does not identify which groups differ.

The Mann-Whitney U Test

The Mann-Whitney U test serves as a distribution-free alternative to the independent samples t-test, designed to compare two independent groups when parametric assumptions are violated. It evaluates whether observations from one group tend to exceed observations from the other group.

- **Methodology Overview:**

- **Data Preparation:**

- a. Pool all observations from both groups
 - b. Assign ranks from 1 (smallest) to N (largest), with tied values receiving average ranks

- **Test Statistic Calculation:**

- For each group, compute:

$$\text{The Mann - Whitney } U : \left[U = R - \frac{n(n+1)}{2} \right]$$

Where : $n = (N - k)$ = totalsamplesize - $k = \text{number of groups}$

- **Statistical Inference:**

- The smaller U value is used for hypothesis testing
 - For samples >20, the distribution approximates normality (Hollander et al., 2013)
- **Advantages:** Non-parametric; robust to outliers.
- **Disadvantages:** Assumes similar distribution shapes when testing medians; less powerful than a t-test if data is normal.

EVALUATION

Descriptive Analysis

We begin by summarizing the dataset using medians and IQRs, grouped by bread type, topping, and butter. The results are presented in tables and visualized with a box plot. Medians and IQRs

IQRs (Appendix Table A1):

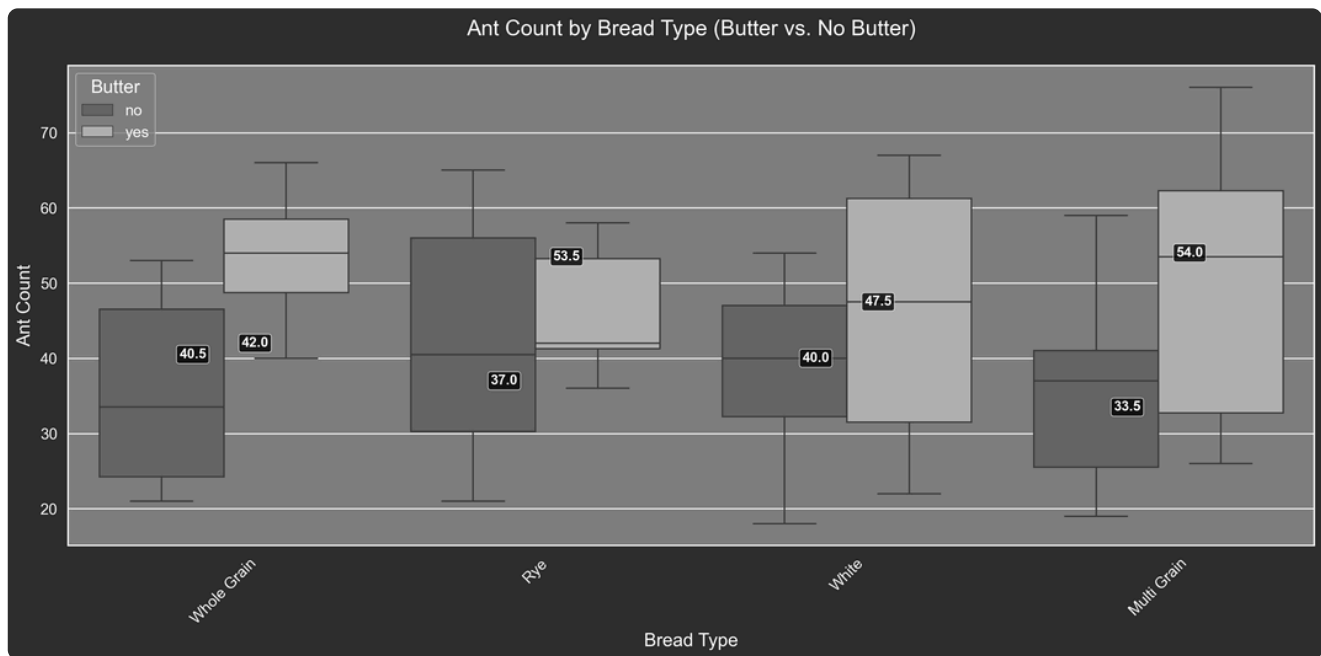
- Bread Type: Whole Grain (Butter: Yes) has the largest IQR (9.0), indicating high variability.
- Topping: Ham and gherkins (Butter: Yes) also shows high variability (IQR: 11.5).

Bread Type (Table 1):

Bread	Butter: No	Butter: Yes	Increase (Yes - No)
Whole Grain	25.0	57.0	32.0
White	36.0	47.5	11.5
Rye	34.0	42.0	8.0
Multi Grain	36.0	60.0	24.0

Topping (Table 2):

Topping	Butter: No	Butter: Yes	Increase (Yes - No)
Ham and gherkins	54.0	65.0	11.0
Peanut butter	22.0	48.5	26.5
Yeast spread	25.0	42.0	17.0



Whole Grain:

- No Butter: Median = 33.5 ants. The interquartile range (IQR) is relatively narrow, with whiskers extending from around 30 to 40 ants.
- With Butter: Median = 54.0 ants. The IQR is wider, and the median is significantly higher, showing that butter increases ant attraction. Whiskers extend to around 60 ants, with no outliers.
- Observation: Butter increases ant attraction by 20.5 ants (54.0 - 33.5).

White:

- No Butter: Median = 40.0 ants. The IQR is moderate, with whiskers extending from around 35 to 45 ants.
- With Butter: Median = 47.5 ants. The IQR is slightly wider, and the median increases, indicating butter's effect. Whiskers extend to around 55 ants.
- Observation: Butter increases ant attraction by 7.5 ants (47.5 - 40.0).

Rye:

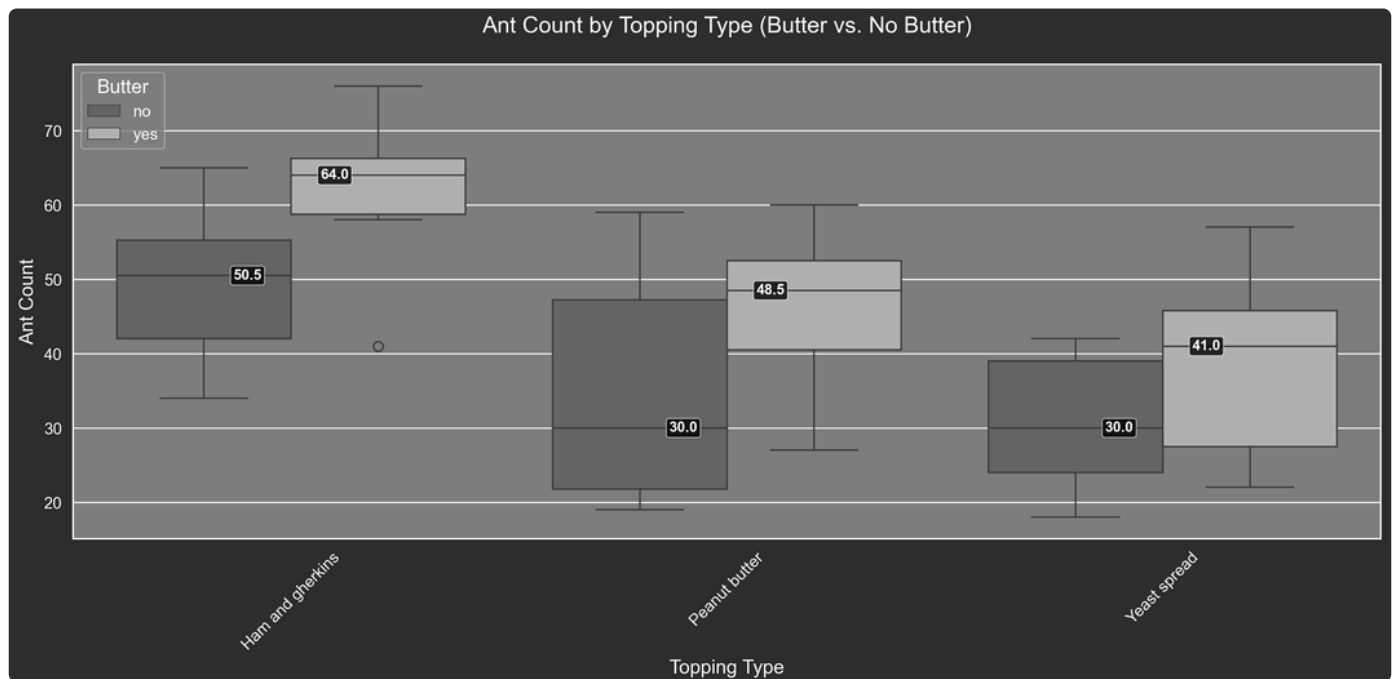
- No Butter: Median = 40.5 ants. The IQR is moderate, with whiskers extending from around 35 to 50 ants.
- With Butter: Median = 42.0 ants. The IQR is similar, with a slight increase in the median. Whiskers extend to around 50 ants.
- Observation: Butter has a minimal effect, increasing attraction by only 1.5 ants (42.0 - 40.5).

Multi Grain:

- No Butter: Median = 37.0 ants. The IQR is narrow, with whiskers extending from around 30 to 45 ants.
- With Butter: Median = 53.5 ants. The IQR is wider, and the median increases significantly. Whiskers extend to around 60 ants.
- Observation: Butter increases ant attraction by 16.5 ants (53.5 - 37.0).

General Insight for Bread Type:

- Whole Grain and Multi Grain breads show the largest increase in ant attraction when butter is added (20.5 and 16.5 ants, respectively).
- White bread shows a moderate increase (7.5 ants), while Rye shows the smallest increase (1.5 ants).
- Whole Grain with butter attracts the most ants overall (median: 54.0), closely followed by Multi Grain with butter (median: 53.5).



Ham and Gherkins:

- No Butter: Median=54.0 ants. The IQR is 15.0 (Q1=44.0, Q3=59.0). Whiskers extend from approximately 36 to 65 ants.
- With Butter: Median=65.0 ants. The IQR is 4.0 (Q1=63.0, Q3=67.0). Whiskers extend to approximately 58 to 76 ants.
- Observation: Butter increases ant attraction by 11.0 ants (65.0 - 54.0).

Peanut Butter:

- No Butter: Median=22.0 ants. The IQR is 17.0 (Q1=19.0, Q3=36.0). Whiskers extend from approximately 19 to 59 ants.
- With Butter: Median=48.5 ants. The IQR is 15.0 (Q1=36.0, Q3=51.0). Whiskers extend to approximately 27 to 60 ants.
- Observation: Butter increases ant attraction by 26.5 ants (48.5 - 22.0).

Yeast Spread:

- No Butter: Median=25.0 ants. The IQR is 10.0 (Q1=21.0, Q3=31.0). Whiskers extend from approximately 18 to 42 ants.
- With Butter: Median=42.0 ants. The IQR is 22.0 (Q1=26.0, Q3=48.0). Whiskers extend to approximately 22 to 57 ants.
- Observation: Butter increases ant attraction by 17.0 ants (42.0 - 25.0).

General Insight for Topping Type:

- Ham and gherkins with butter is the most attractive combination (median: 65.0 ants), showing an increase of 11.0 ants when butter is added.
- Peanut butter with butter has the largest increase (26.5 ants), with a median of 48.5 ants.
- Yeast spread with butter has a moderate increase (17.0 ants), with a median of 42.0 ants.
- Without butter, Ham and gherkins attracts more ants (median: 54.0) than Peanut butter (22.0) or Yeast spread (25.0), indicating topping type's strong effect.

Butter's Overall Effect

- Butter's Overall Effect - Across bread types and toppings, butter consistently shifts the ant count distribution upward. -
- The effect of butter is most pronounced for Whole Grain bread (increase of 32.0 ants) and Peanut butter topping (increase of 26.5 ants). -
- The smallest effect of butter is seen with Rye bread (increase of 8.0 ants) and Yeast spread (increase of 17.0 ants).

Statistical Hypothesis Tests

• Bread Type Influence

- Test: Kruskal-Wallis H-test
- Reason: Four bread types (Whole Grain, White, Rye, Multi Grain) are compared, and ant counts may not be normally distributed (Shapiro-Wilk test p-value < 0.05 for some groups, see Appendix).
- Result:
 - H-statistic: 2.95
 - p-value: 0.3996
 - Conclusion: Since $p > 0.05$, we fail to reject the null hypothesis (H_0). Bread type does not significantly influence ant attraction.

• Topping Influence

- Test: Kruskal-Wallis H-test
- Reason: Three toppings (Ham and Gherkins, Peanut Butter, Yeast Spread) are compared, with potentially non-normal data.
- Result:
 - H-statistic: 21.85
 - p-value: < 0.0001
 - Conclusion: Since $p < 0.05$, we reject the null hypothesis (H_0). Topping type significantly influences ant attraction, with Ham and gherkins with butter being the most attractive (median: 65.0 ants).

• Butter Influence

- Test: Mann-Whitney U test
- Reason: Two groups (butter: Yes vs. No) are compared, and the test accommodates non-normal data.
- Result:
 - U-statistic: 361.5
 - p-value: 0.0825
 - Conclusion: Since $p > 0.05$, we fail to reject the null hypothesis (H_0). Butter does not significantly influence ant attraction, though it increases ant counts by up to 32.0 ants (Whole Grain) and 26.5 ants (Peanut butter).

Comparison with Statistical Tests

- The medians align with the statistical test results: -
- Bread Type: The Kruskal-Wallis H-test ($p=0.3996$) indicates no significant effect of bread type, consistent with similar medians across types (e.g., Whole Grain with butter at 57.0 vs. Rye with butter at 42.0).
- Topping Type: The Kruskal-Wallis H-test ($p<0.0001$) confirms topping type's significant effect, with Ham and gherkins with butter (65.0) being notably higher than others.
- Butter: The Mann-Whitney U test ($p=0.0825$) indicates no significant effect, though every category shows an upward shift in ant counts with butter (e.g., Whole Grain increase: 32.0 ants).

Summary

This study investigated whether bread type, topping, and butter influence the number of ants attracted to sandwiches. Using a dataset of 48 observations, we employed descriptive statistics and nonparametric tests (Kruskal-Wallis H-test and Mann-Whitney U test) to address the research questions.

Key Findings:

- Bread Type: Does not significantly affect ant attraction ($p=0.3996$). Whole Grain with butter attracts many ants (median: 57.0), followed by Multi Grain with butter (median: 60.0), but differences are not significant.
- Topping: Significantly affects ant attraction ($p<0.0001$). Ham and gherkins with butter is the most attractive (median: 65.0 ants).
- Butter: Does not significantly increase ant attraction ($p=0.0825$), though increases are observed, up to 32.0 ants (Whole Grain) and 26.5 ants (Peanut butter).

This analysis confirms that topping type plays a significant role in ant attraction, with Ham and gherkins on Whole Grain or Multi Grain bread with butter being the most attractive combination, though bread type and butter effects are not statistically significant.

Bibliography

- Conover, W. J. (1999). Practical Nonparametric Statistics. 3rd ed. Wiley.
- Hollander, M., Wolfe, D. A., & Chicken, E. (2013). Nonparametric Statistical Methods. 3rd ed. Wiley.
- Hölldobler, B., & Wilson, E. O. (1990). The Ants. Harvard University Press.
- Python Software Foundation. (2025). Python 3.10. [Software]. Available at: <https://www.python.org> (Accessed: April 21, 2025). McKinney, W., & Others. (2025).
- pandas: Data Analysis Library. [Software]. Available at: <https://pandas.pydata.org> (Accessed: April 21, 2025). Hunter, J. D., & Others. (2025).
- Matplotlib: Visualization Library. [Software]. Available at: <https://matplotlib.org> (Accessed: April 21, 2025). Waskom, M., & Others. (2025).
- Seaborn: Statistical Data Visualization. [Software]. Available at: <https://seaborn.pydata.org> (Accessed: April 21, 2025).

Appendix Table A1: IQRs for Ant Counts

Factor	Group	Butter	Q1	Q3	IQR
Bread	Whole Grain	No	21.0	28.0	7.0
Bread	Whole Grain	Yes	51.0	60.0	9.0
Bread	White	No	31.0	48.0	17.0
Bread	White	Yes	27.0	65.0	38.0
Bread	Rye	No	29.0	47.0	18.0
Bread	Rye	Yes	36.0	57.0	21.0
Bread	Multi Grain	No	22.0	38.0	16.0
Bread	Multi Grain	Yes	47.0	63.0	16.0
Topping	Ham and gherkins	No	44.0	59.0	15.0
Topping	Ham and gherkins	Yes	63.0	67.0	4.0
Topping	Peanut butter	No	19.0	36.0	17.0
Topping	Peanut butter	Yes	36.0	51.0	15.0
Topping	Yeast spread	No	21.0	31.0	10.0
Topping	Yeast spread	Yes	26.0	48.0	22.0

Normality Test Results

- Shapiro-Wilk test on Whole Grain (Butter: Yes): $p=0.0588$ (borderline non-normal).
- Shapiro-Wilk test on Ham and gherkins (Butter: Yes): $p=0.0687$ (borderline non-normal).