TU DORTMUND UNIVERSITY

Introductory Case Studies

Project 2: Comparison of K distributions

Ants on the Picnic Blanket

Lecturers:

Prof. Dr. Paul Bürkner
Dr. Daniel Habermann
Lars Kühmichel
Svenja Jedhoff

Author: Sourav Poddar

Matriculation No: 236625

Group Number: 11

Group Members:
Dia Tasneem Naser
Apu Kumar Saha
Nandita Chakrobortty

November 11, 2025

Contents

1 Introduction

Ants are one of the most common and important insects for the environment. They help ecosystems work by aerating the soil, cycling nutrients, and spreading seeds.Scientists have been interested in their amazing ability to find and use food sources for a long time

In this study, we investigated whether different sandwich compositions exhibit varying levels of attractiveness to ants. Our analysis uses the `Sandwich` dataset (Mackisack 1994), which contains 48 experimental observations from the sequential placement of 24 distinct sandwich types near an ant hill. The dataset records quantitative counts of ants and categorical descriptors of bread type, sandwich toppings, and butter presence.

Understanding how difficult it may be to compare several variables at once, we use a multiple-test problem technique to fully investigate the impact of the aforementioned characteristics.In order to learn more about the structure and contents of the dataset, we initially performed exploratory data analysis, or EDA. We next carried out statistical tests, such as ANOVA, to look into how different factors affected ant attraction. We conducted Levene's test and the Shapiro-Wilk test on the variables as prerequisites to determine whether or not the ANOVA approach is appropriate in this case. The significant differences found by the ANOVA tests were then further examined using Tukey's Honestly Significant Difference (HSD) test. We can find particular differences between the groups by using Tukey's HSD test, which permits pairwise comparisons between group means. Ultimately, we deduced the elements controlling ant behavior by interpreting the findings of these investigations.

A summary of the dataset and the project's goal are provided in Section 2. The statistical methods are explained in Section 3. In Section 4, the statistical anlysis are illustrated. This section also compares the outcomes of multiple testing. An overview of the entire project and possible limitation are shown in the Sections 5.

2 Problem Statement

2.1 Project Objectives

We aim to explore which factors influence ant attraction to sandwiches placed on a picnic blanket. Specifically, we test whether the **type of bread (Brot)**, the **type of topping (Belag)**, or the **presence of butter (Butter)** significantly affects how many ants are attracted. To answer these questions, we employ a combination of descriptive statistics, assumption checks, and both one-way and two-way Analysis of Variance (ANOVA) to properly test these relationships and draw meaningful conclusions from our data.

2.2 Data Description

For our analysis, we worked with the dataset **Sandwich**. Each observation(n=48) recorded the specific conditions of a sandwich and the resulting ant count. The variables we examined were:

- Ameisen (Number of Ants): This was our dependent variable, recorded as a continuous integer.
- Brot (Bread Type): This categorical independent variable had four levels that we compared: Mehrkorn (multigrain), Roggen (rye), Vollkorn (whole grain), and Weiß (white).
- Belag (Topping): This was another categorical independent variable with three levels: Erdnussbutter (peanut butter), Hefe-Brotaufstrich (yeast spread), and Schinken-Essiggurken (ham and pickles).
- Butter: Our final (binary)categorical independent variable had two levels: ja (yes) and nein (no), indicating whether butter was applied.

After conducting preliminary data screening, we confirmed that we had a complete dataset with no missing values across any of the variables we planned to analyze.

Table 1: Data Structure Overview

Variable	**Type**	**Levels**	**Observations per Level**
Ameisen	Continuous	–	48 total
Brot	Categorical	4	12 each
Belag	Categorical	3	16 each
Butter	Categorical	2	24 each

3 Statistical Methods

Prior to analysis, we performed comprehensive data quality checks including verification of missing values and assessment of variable types. Categorical variables (Brot, Belag, Butter) were converted to factors to ensure proper treatment in statistical models.We calculated measures of central tendency (mean, median) and dispersion (standard deviation, interquartile range) for each factor level.We also generated boxplots to visualize distribution patterns

3.1 Hypothesis Test

A statistical test is a way to figure out if a hypothesis regarding an unknown parameter in the distribution of a random variable should be accepted or rejected. There are two types of hypotheses: the null hypothesis, which is written as H_0 , and the alternative hypothesis, which is written as H_A. If H_0 is accepted, H_A is rejected, and if H_A is accepted, H_0 is rejected (Neyman and Pearson 1933).

There are two kinds of errors in hypothesis testing:

Type I error (false positive): H_0 is true but is rejected. The *significance level* α is the pre–specified probability of a Type I error (commonly $\alpha = 0.05$ or 0.01). A smaller α requires stronger evidence to reject H_0. **Type II error** (false negative): H_0 is false but is not rejected. Its probability is denoted β, and the *power* of the test is $1 - \beta$. For fixed data, decreasing α generally increases β, and vice versa; increasing the sample size n typically reduces both error probabilities by increasing power.

A common decision rule compares the p-value with α. The p-value (observed significance level) is the probability, computed under H_0, of obtaining a test statistic at least as extreme as the one observed. If the p-value is less than the chosen level α, reject H_0 (Lehmann and Romano 2005)

3.2 Normality Assessment

- **Shapiro-Wilk Test:**

  To verify the critical assumption of normality for parametric ANOVA tests, we employed a multi-faceted approach combining formal statistical tests and graphical methods. The Shapiro–Wilk test evaluates the null hypothesis that a sample comes from a normally distributed population. The test statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i \, x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

  where: $x_{(i)}$ are the ordered sample values (statistical order), a_i are the coefficients obtained from the regression, $\bar{x}$ is the sample mean,n is the sample size.

  The statistic W takes values in $(0, 1]$, with values closer to 1 indicating stronger evidence of normality (smaller values suggest departures from normality)(Shapiro and Wilk 1965).

- **Q-Q Plots:** We performed visual inspection of quantile-quantile plots to assess deviations from normality.

3.3 Levene's Test

We employed this test using median as center for robustness to test equality of variances across groups. It used to check if our data sets fulfill the homogeneity of variance assumption (Levene 1960) before we performed the Analysis of Variance (ANOVA).

Hypotheses:

$$H_0:\ \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_g^2 \qquad \text{vs} \qquad H_A:\ \text{not all } \sigma_i^2 \text{ are equal.}$$

Levene's test statistic, W is calculated as:

$$W \;=\; \frac{(N-k)}{(k-1)} \cdot \frac{\displaystyle\sum_{i=1}^{k} n_i \left(\bar{z}_i - \bar{z}\right)^2}{\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2},$$

where: N is the total number of observations, k is the number of groups, n_i is the number of observations in the i-th group, $\bar{z}_i$ is the mean of the i-th group, $\bar{z}$ is the overall mean, z_{ij} is the absolute deviation from the group center

$$W > F_{\alpha,\, k-1,\, N-k}$$

The statistic W follows an F distribution with $k-1$ and $N-k$ degrees of freedom Under the null hypothesis of equal variances across groups (Levene, 1960).

3.4 One-Way Analysis of Variance (ANOVA)

The statistical model for a one-way ANOVA with k groups is

$$Y_{ij} \;=\; \mu \;+\; \alpha_j \;+\; \varepsilon_{ij},$$

where

- Y_{ij} is the observation (e.g., number of ants) in the j-th group and i-th replicate, μ is the overall population mean, $\alpha_j = \mu_j - \mu$ is the effect of the j-th group, with μ_j the mean of group j, ε_{ij} is the random error term, with $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

The hypothesis test is stated as

- **Null Hypothesis (H_0):** $\mu_1 = \mu_2 = \cdots = \mu_k$ (all group means are equal).
- **Alternative Hypothesis (H_A):** At least one μ_j is different from the others.

We conducted separate one-way ANOVAs for each factor (Brot, Belag, Butter) to test the global null hypothesis of no difference between group means.

3.5 Two-Way Analysis of Variance (ANOVA)

For interaction analysis, we specified the two-way ANOVA model, also known as two-factor ANOVA. Two-way ANOVA is a statistical test used to examine the influence of two categorical independent variables on a continuous dependent variable (Montgomery 2019). It assesses whether there are any statistically significant interactions between the two independent variables and whether they have a significant effect on the dependent variable.

$$Y_{ijk} \;=\; \mu \;+\; \alpha_i \;+\; \beta_j \;+\; (\alpha\beta)_{ij} \;+\; \varepsilon_{ijk},$$

where

- α_i is the effect of the i-th level of the first factor (Factor A), β_j is the effect of the j-th level of the second factor (Factor B), $(\alpha\beta)_{ij}$ is the interaction effect between Factors A and B at levels i and j, ε_{ijk} is the random error term for replicate k in cell (i, j),typically assumed i.i.d. $\mathcal{N}(0, \sigma^2)$.

3.6 Multiple Comparisons Procedure

3.6.1 The Multiple Testing Problem

When conducting multiple simultaneous hypothesis tests, we recognized that the family-wise error rate (FWER) increases substantially. For k independent tests at significance level α, the probability of at least one Type I error is

$$\text{FWER} \;=\; 1 - (1 - \alpha)^k.$$

This inflation necessitates appropriate correction procedures to maintain the overall Type I error rate.

3.6.2 Tukey's Honestly Significant Difference (HSD)

To control the FWER, we use Tukey's Honest Significant Difference (HSD) test (Tukey 1977). This test is a single-step multiple comparison procedure that applies a correction to the critical value based on the Studentized range distribution, allowing us to simultaneously compare all possible pairs of means while keeping the FWER at the desired level (e.g., 0.05).

$$\text{HSD} \;=\; q_{\alpha,\,k,\,df} \;\times\; \sqrt{\frac{MS_{\text{error}}}{n}}$$

where

- $q_{\alpha,\,k,\,df}$ is the studentized range statistic,

- k is the number of groups,
- df is the degrees of freedom for the error term,
- MS_{error} is the mean square error from ANOVA,
- n is the sample size per group.

3.7 Tools

We performed the analysis using R (version 4.4.3) (R Core Team 2024). Key packages included *dplyr* for data manipulation, *ggplot2* for visualization, *car* for Levene's Test (Fox and Weisberg 2019), and the base R function *TukeyHSD()* for post-hoc analysis.

4 Statistical analysis

4.1 Descriptive comparisons

An initial exploratory analysis was conducted to understand the distribution of the number of ants *(Ameisen)* across the different levels of the categorical factors: bread type *(Brot)*, topping *(Belag)*, and the presence of butter *(Butter)*.

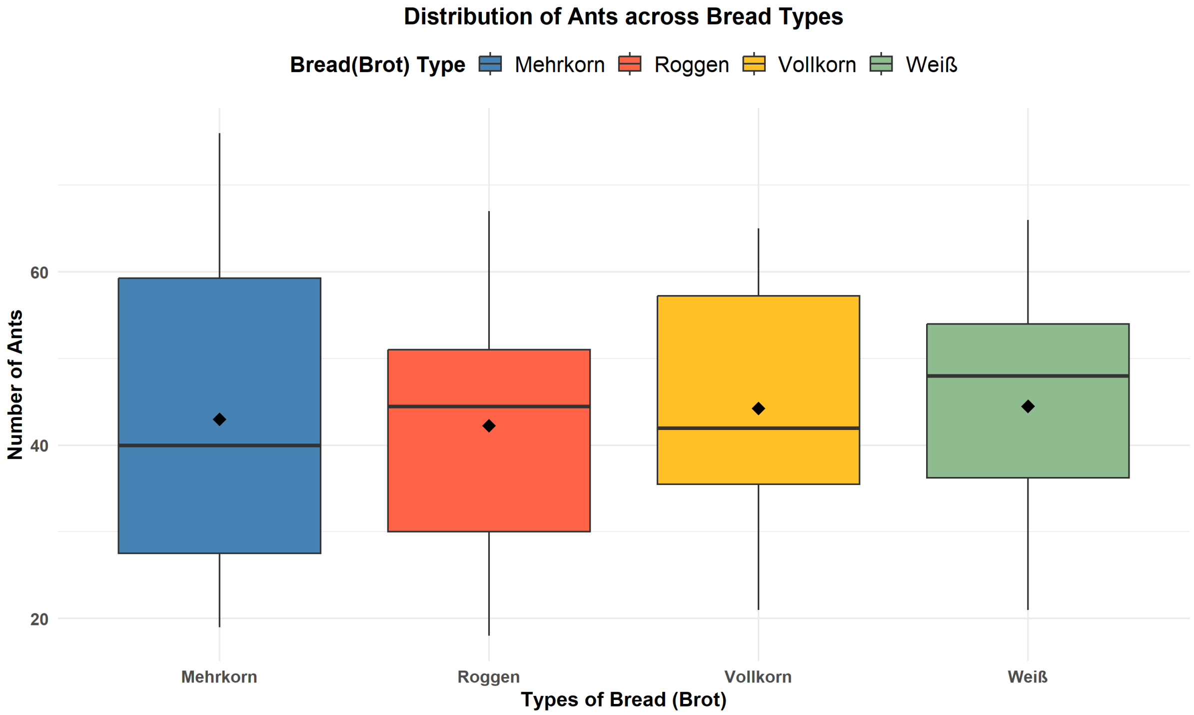

Figure 1: Distribution of Ants across Bread Types

From the Figure 1 and Table 2 presents the summary statistics of the *Brot* variable; the totals for each bread type are relatively consistent.

Table 2: Descriptive Statistics for Number of Ants by Bread Type *Brot*

Brot	**sum**	**max**	**min**	**mean**	**median**	**std**
Vollkorn	531	65	21	44.25	42	13.40
Mehrkorn	516	76	19	43.00	40	18.23
Roggen	507	67	18	42.25	44.5	15.86
Weißbrot	534	66	21	44.50	48	14.61

In Table 3, the topping *Schinken–Essiggurken* was the most frequently occurring also has the highest mean ant count (55.5)in the *Belag* variable, while *Hefe–Brotaufstrich* had the lowest mean (34.63).

Table 3: Descriptive Statistics for Number of Ants by (*Belag*)

Belag	**sum**	**max**	**min**	**mean**	**median**	**std**
Erdnussbutter	646	60	19	40.38	44.5	14.18
Hefe-Brotaufstrich	554	57	18	34.63	34.5	11.16
Schinken-Essiggurken	888	76	34	55.50	58.5	12.06

According to Table 4 presents a statistical summary regarding the presence of butter in sandwiches, where Sandwiches with butter have higher means than those without (48.9 vs. 38.1).

Table 4: Descriptive Statistics for Number of Ants by Butter Presence (*Butter*)

Butter	**n**	**Mean**	**Std. Dev.**	**Median**	**IQR**
ja	24	48.9	14.5	49.0	18.5
nein	24	38.1	14.1	37.0	23.2

4.2 Assumption Checking

Before performing the ANOVA, we checked the assumptions of normality of residuals and homogeneity of variances.

4.2.1 Normality of Residuals:

The Shapiro-Wilk test was performed on the residuals of the ANOVA models for each factor.As shown in Table 5, the p-values for the residuals of all three models (Brot, Belag, and Butter) are greater than the significance level $\alpha = 0.05$. This suggests that we cannot reject the null hypothesis that the residuals are normally distributed. Group-wise QQ plots by factor show no serious deviations. The Q-Q plots in Appendix Figure 6 support this finding.

Table 5: Shapiro–Wilk Test Results for Normality of Residuals

Factor	**Model**	**W Statistic**	***p*-value**	**Assumption Met ($\alpha = 0.05$)**
Brot	Ameisen $\sim$ Brot	0.96779	0.20750	Yes
Belag	Ameisen $\sim$ Belag	0.95688	0.07552	Yes
Butter	Ameisen $\sim$ Butter	0.97520	0.39790	Yes

4.2.2 Homogeneity of Variances:

We used Levene's Test (center = median) to check the homogeneity of variances across the groups for each factor. Table 6 shows that all P-values were non-significant ($p > 0.05$) (Wasserstein and Lazar 2016), confirming that the assumption of homogeneity of variances was met for all factors.

Table 6: Levene's Test Results for Homogeneity of Variances

Factor	**Degrees of Freedom (Df)**	**F value**	***p*-value**	**Assumption Met ($\alpha = 0.05$)**
Brot	3, 44	0.5266	0.6663	Yes
Belag	2, 45	0.6345	0.5349	Yes
Butter	1, 46	0.0014	0.9702	Yes

Since both the normality and homogeneity of variances assumptions were satisfied, it is appropriate to proceed with a parametric one-way ANOVA.

4.3 One-Way ANOVA

We conducted three separate One-Way ANOVA to assess the influence of each factor on the number of ants attracted (Ameisen)(Welch 1951). The summary results are presented in Table 7

Table 7: Summary of One-Way ANOVA Tests for All Factors

Factor	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr($> F$)**	**Significance**
Brot	3	40	13.5	0.055	0.983	Not Significant
Belag	2	3720	1860	11.85	7.36e-05	Highly Significant
Butter	1	1387	1386.7	6.787	0.0123	Significant

Effect of Bread (Brot):

The ANOVA for the bread type was not statistically significant ($F(3, 44) = 0.055, p = 0.983$). We therefore conclude that there is no evidence that the type of bread affects the number of ants attracted. This is visually supported by the descriptive boxplot shown in Figure 1, where the mean values (black dots) are very similar across all four bread types.

Effect of Topping (Belag):

The ANOVA for the topping type was highly statistically significant ($F(2, 45) = 11.85, p < 0.001$). This result allows us to reject the null hypothesis and conclude that the type of topping significantly influences ant attraction.

Effect of Butter(Butter):

The ANOVA for the presence of butter was statistically significant ($F(1, 46) = 6.787, p = 0.0123$). We conclude that the presence of butter significantly increases the number of attracted ants.

In Figure 2 shows the descriptive distribution, confirming that the mean number of ants for sandwiches with butter ("ja") is higher than those without ("nein").

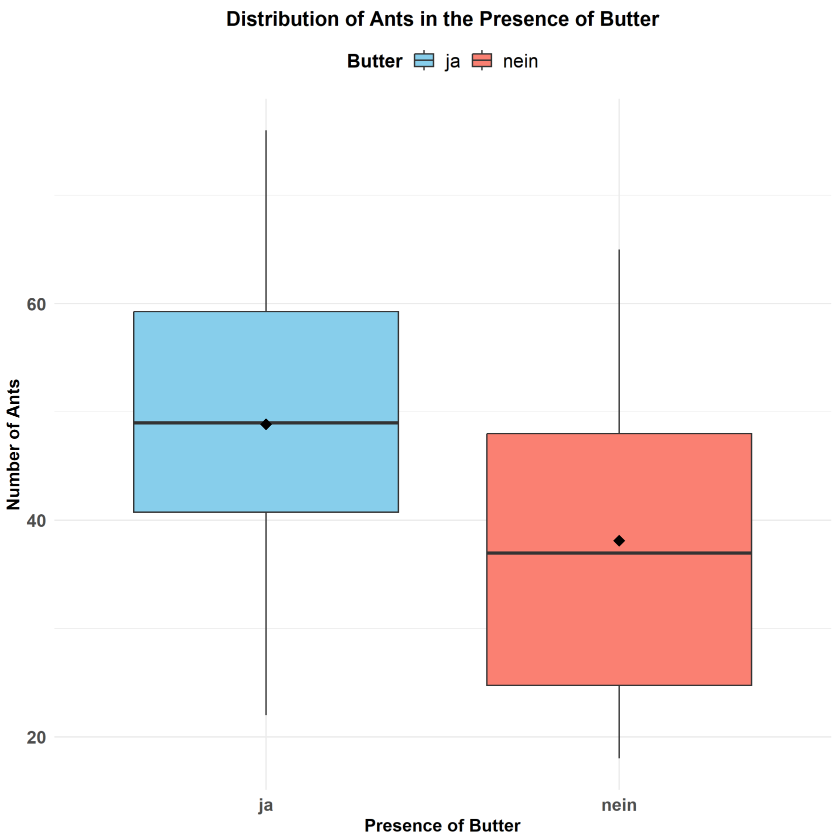

Figure 2: Distribution of Ants in the Presence of Butter

4.4 Multiple comparisons (Tukey HSD)

Since the ANOVAs for Topping (Belag) and Butter were significant, post-hoc pairwise comparisons were necessary to identify which specific groups differ (Tukey 1949). Conducting multiple statistical tests increases the family-wise error rate (FWER)—the probability of making at least one Type I error (false positive) across all tests. This is known as the multiple testing problem.(Hochberg and Tamhane 1987)

Pairwise Comparisons for Topping (Belag): We examine the descriptive statistics and distribution in Figure 3, which clearly shows that the mean number of ants for the "Schinken-Essiggurken" topping is substantially higher than other 2 toppings.

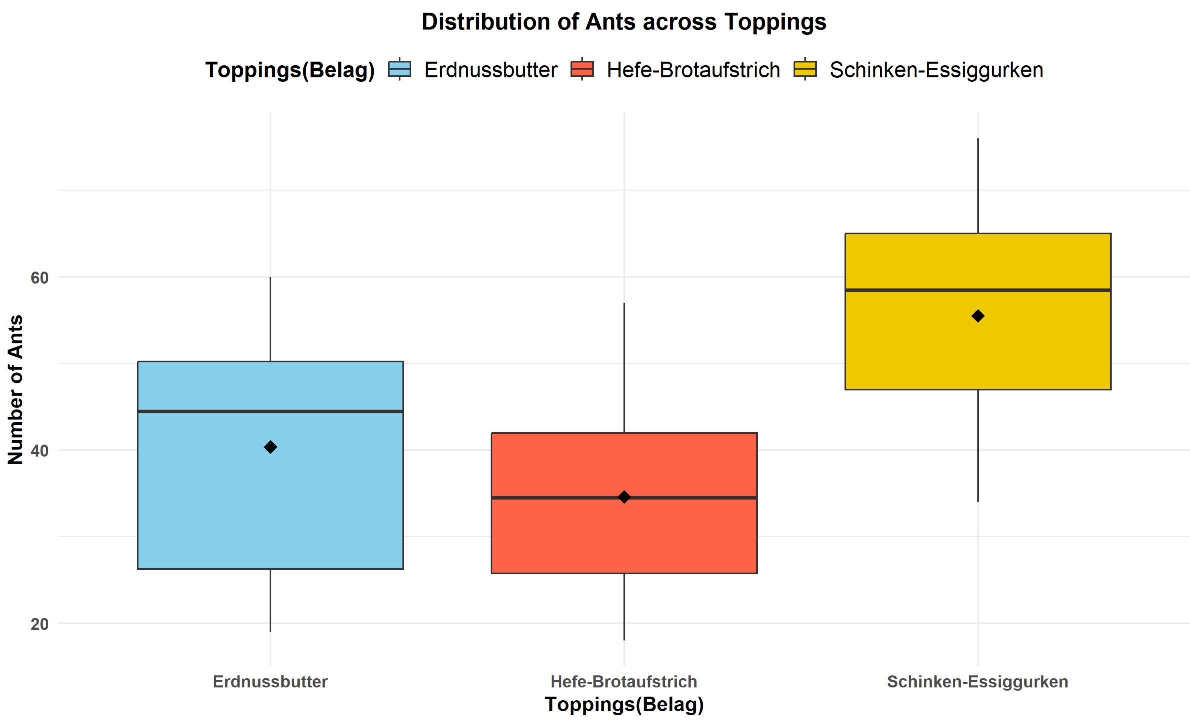

Figure 3: Distribution of Ants across Toppings

To determine which specific pairs are different, we conducted a post-hoc analysis using Tukey's Honest Significant Difference (HSD) test to control the Family-Wise Error Rate (FWER) at $\alpha = 0.05$.

Table 8: Tukey's HSD for topping (*Belag*)

Comparison	**Diff)**	**95% Lwr**	**95% Upr**	**p-adj**
Schinken-Essiggurken vs. Hefe-Brotaufs	20.875	10.138	31.612	0.00007
Schinken-Essiggurken vs. Erdnussbutter	15.125	4.388	25.862	0.00383
Hefe-Brotaufs vs. Erdnussbutter	-5.750	-16.487	4.987	0.40374

From the Table 8, we can say that the "Schinken-Essiggurken" (ham pickles) topping is significantly more attractive to ants than both "Erdnussbutter" (peanut butter) and "Hefe-Brotaufstrich" (yeast spread). There is no significant difference between them.

Pairwise Comparison for Butter (Butter):

Table 9: Tukey HSD Results for Butter (*Butter*)

Comparison	Mean Difference	Adjusted p-value	Significant
nein – ja	-10.75	0.012	**Yes**

The analysis shows that sandwiches **without butter** (`nein`) attract **significantly fewer ants** than sandwiches **with butter** (`ja`). The mean difference between the two groups is-10.75, and the p-value associated with this difference is 0.0123, indicating statistical significance. Therefore, it can be inferred that the presence of butter on the sandwich influences ant attraction

4.5 Two-Way ANOVA

As an additional check on the independence of the factor effects, we performed two-way ANOVAs to test for potential interactions between factors.

1. **Topping (*Belag*)** $\times$ **Butter.** The interaction between these two factors was found to be non-significant $F(2, 42) = 0.209,\ p = 0.812$. This means the effect of butter is generally consistent across all topping types. Figure 4 visually represents this parallel effect.

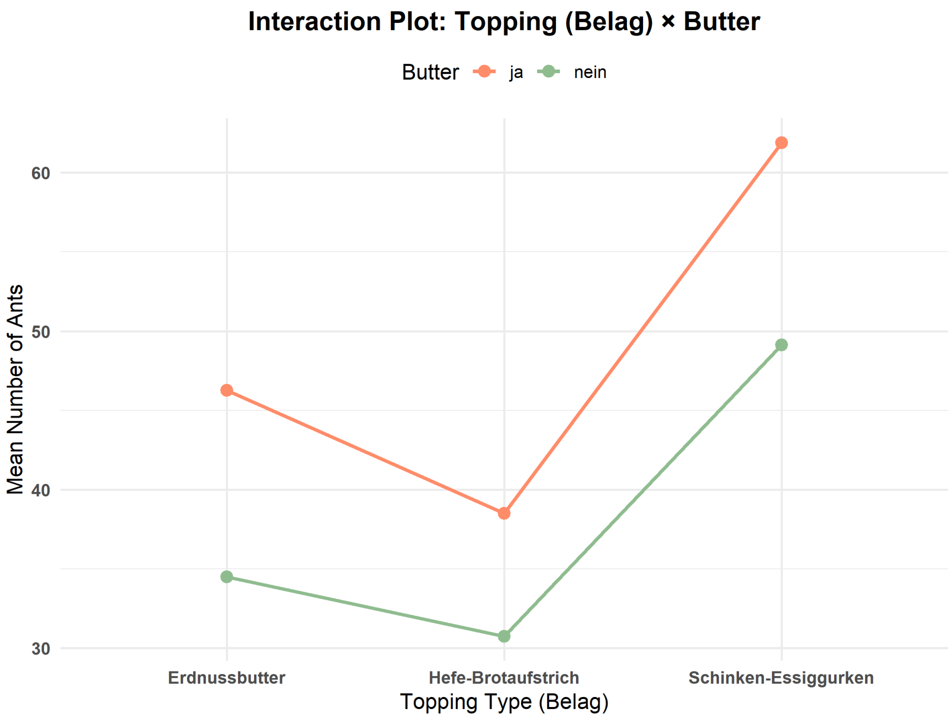

Figure 4: Interaction Plot: Topping (Belag) $\times$ Butter

2. **Bread (*Brot*) × Butter.** The interaction between the bread type and the presence of butter was also non-significant $F(3, 40) = 0.562$, $p = 0.643$. This indicates that the effect of butter does not depend on the type of bread used. Figure 5 shows this result.

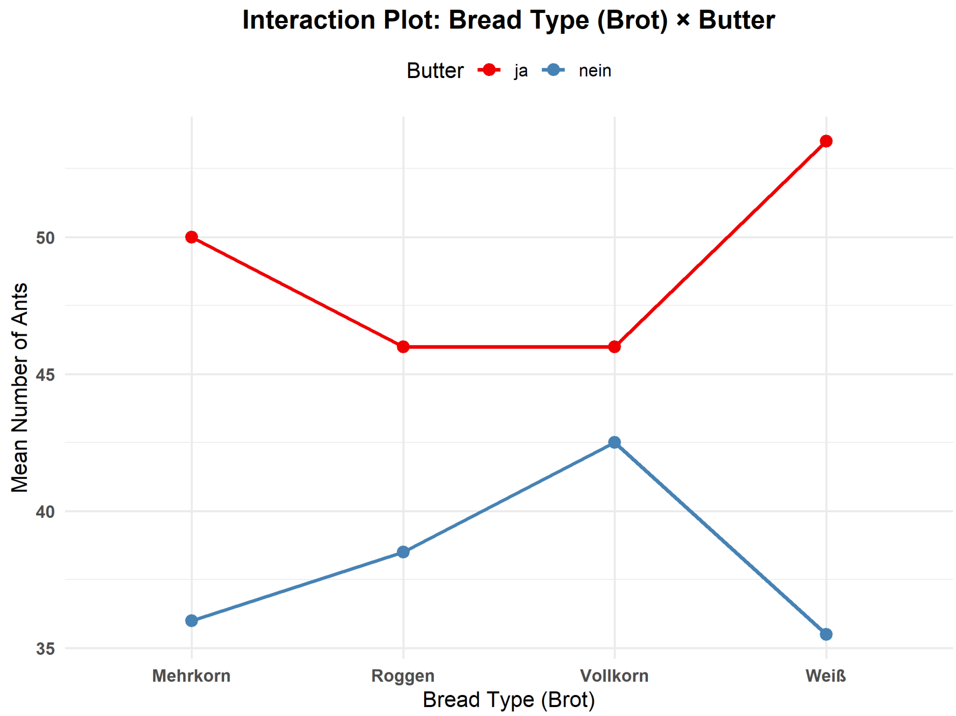

Figure 5: Interaction Plot: Bread Type (Brot) × Butter

Table 10: ANOVA table for Belag and Butter

Variable	**sum_sq**	**df**	**F**	**PR($> F$)**
Belag	3720.50	2.0	13.895	0.000023
Butter	1386.75	1.0	10.358	0.002486
(Belag):(Butter)	56.00	2.0	0.209	0.812113
Residual	5622.75	42.0	NaN	NaN

Table 11: ANOVA table for Brot and Butter

Variable	**sum_sq**	**df**	**F**	**PR($> F$)**
Brot	40.50	3.0	0.060	0.980
Butter	1386.75	1.0	6.17	0.017
(Brot):(Butter)	378.75	3.0	0.562	0.643
Residual	8980.00	40.0	NaN	NaN

From Table 11 p-value(0.643) is greater than the significance level (typically 0.05),fail to reject the null hypothesis of no interaction effect.

5 Summary

This study investigates the factors affecting ant attraction to sandwiches using "sandwich.sav" dataset. Through exploratory data analysis and rigorous statistical testing, including ANOVA and post-hoc Tukey's HSD tests, the research identifies key influencers. The analysis, which confirmed underlying statistical assumptions,such as the homogeneity of variances are tested using Levene's test, and normality assumptions are assessed using the Shapiro-Wilk test, determined that the type of bread is irrelevant to ants.

In contrast, the sandwich topping(Belag) is a major factor, with Schinken-Essiggurken proving significantly more attractive than other toppings. The presence of butter was also found to be a consistent and significant attractant. These results provide clear evidence that ant behavior is directly influenced by specific sandwich ingredients.

These results are from a controlled experiment and may vary with different ant species or environmental conditions. Nonetheless, the core finding is clear: to minimize ants, avoid using butter and choose toppings like peanut butter or yeast spread over ham and pickles.

References

Fox, John and Sanford Weisberg (2019). *An R Companion to Applied Regression*. 3rd ed. Thousand Oaks, CA: Sage.

Hochberg, Y. and A.C. Tamhane (1987). *Multiple Comparison Procedures*. New York: Wiley.

Lehmann, Erich L. and Joseph P. Romano (2005). *Testing Statistical Hypotheses*. 3rd ed. New York: Springer.

Levene, Howard (1960). "Robust tests for equality of variances". In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Ed. by Ingram Olkin, S. G. Ghurye, Wassily Hoeffding, William G. Madow, and Henry B. Mann. Stanford University Press, pp. 278–292.

Mackisack, Margaret (1994). "What Is the Use of Experiments Conducted By Statistics Students?" In: *Journal of Statistics Education* 2.1, pp. 1–15. URL: `https://jse.amstat.org/v2n1/mackisack.html`.

Montgomery, Douglas C. (2019). *Design and Analysis of Experiments*. 10th ed. John Wiley & Sons.

Neyman, Jerzy and Egon S. Pearson (1933). "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A* 231, pp. 289–337.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Shapiro, Samuel S. and Martin B. Wilk (1965). "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3-4, pp. 591–611.

Tukey, John W. (1949). "Comparing individual means in the analysis of variance". In: *Biometrics* 5.2, pp. 99–114.

— (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Wasserstein, Ronald L. and Nicole A. Lazar (2016). "The ASA's Statement on p-Values: Context, Process, and Purpose". In: *The American Statistician* 70.2, pp. 129–133.

Welch, Bernard L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach". In: *Biometrika* 38.3-4, pp. 330–336.

A Appendix

A.1 Q-Q Plots

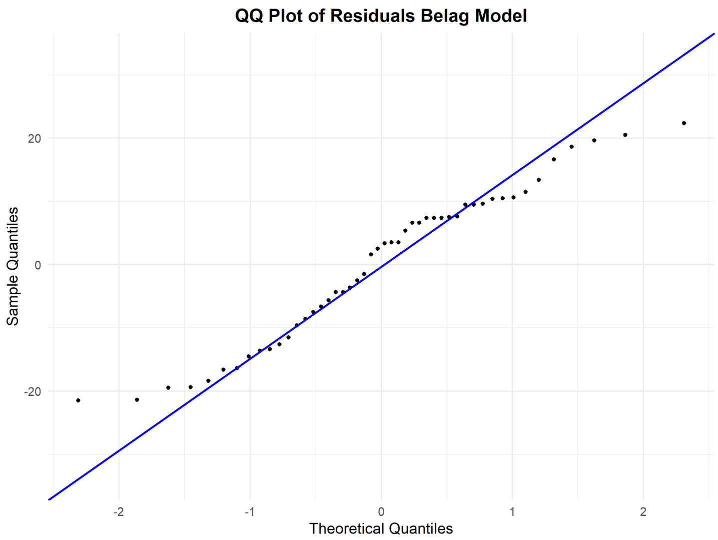

Figure 6: Residual Q-Q Belag

Figure 7: Residual Q-Q Brot

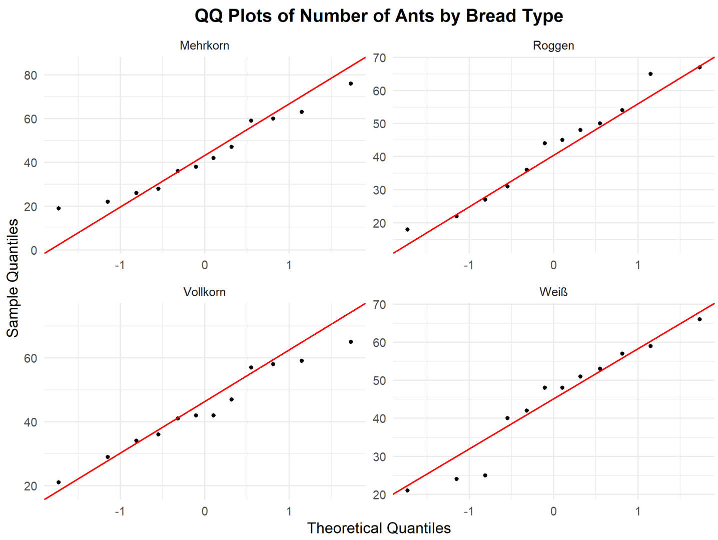

Figure 8: QQ Plots for Group Wise Normality Brot

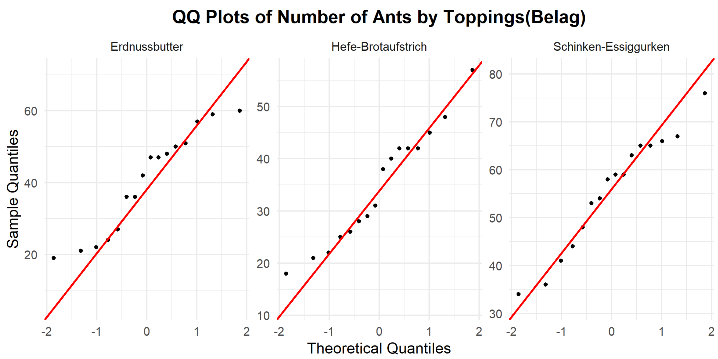

Figure 9: QQ Plots for Group Wise Normality Belag

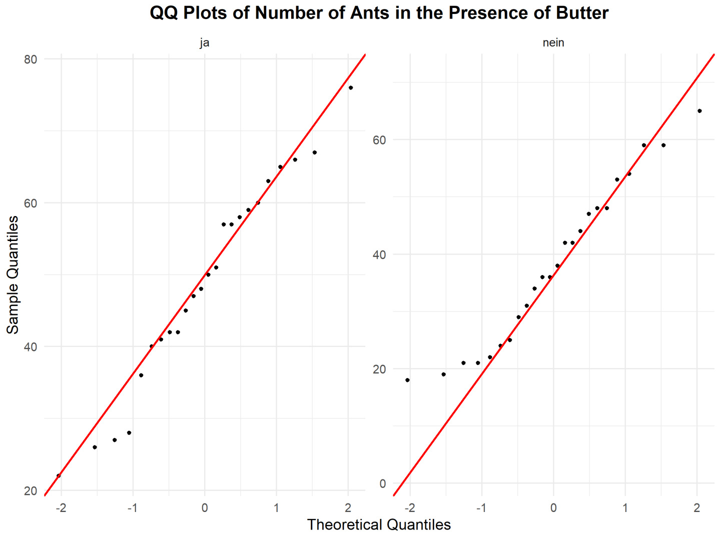

Figure 10: QQ Plots for Group Wise Normality Butter