# Influence of Sandwich Composition on Ant Attraction

Simran Amesar
June 14, 2025

# 1. Introduction

Understanding factors that influence insect behavior, such as ant attraction to food sources, holds practical significance in various fields, from pest control to ecological studies. Ant behavior in response to food types is not only of ecological interest but also illustrates how scent, sugar, and fat content affect animal attraction. This report addresses a curious student's inquiry into whether specific sandwich ingredients attract ants more than others. The student conducted a controlled experiment to investigate the impact of bread type, sandwich topping, and the presence or absence of butter on the number of ants attracted to sandwich pieces.

The primary objective of this report is to conduct a comprehensive statistical analysis of the collected data to determine:

1. Do the type of bread, the topping, and butter have a significant influence on the amount of attracted ants?
2. Which type of bread and which topping attract the most ants?

This report will begin by detailing the experimental setup and the data collected. Subsequently, it will describe the statistical methodologies employed for both descriptive and inferential analyses. The evaluation section will present the results, followed by a summary, interpretation, and discussion of the findings, including limitations and avenues for future research.

# 2. Description of the Problem and Dataset

The experiment aimed to investigate the influence of three independent variables - bread type, topping, and butter presence - on a dependent variable: the number of attracted ants. The study involved placing 24 different sandwich configurations near an ant hill, one after another, in a randomized order to minimize bias from temporal effects or ant foraging patterns. After a predetermined exposure time, a glass bell was placed over each sandwich piece, and the attracted ants were counted.

Each sandwich has a unique combination of:

**Bread Type:** Wholemeal, Multigrain, Rye, White (Four categories)

**Topping:** Ham & Gherkins, Peanut Butter, Yeast Spread (Three categories)

**Butter**: Yes or No (Two categories)

$$4 \times 3 \times 2 \times 2 \text{ (repetition)} = 48 \text{ observations}$$

The data consists of 48 observations.

The dataset, sourced from the student's experiment, consists of 48 rows and 4 columns:

**antCount**: Quantitative variable, representing the number of attracted ants.

**bread**: Categorical variable, representing the type of bread.

**topping**: Categorical variable, representing the sandwich topping.

**butter**: Categorical variable, indicating the presence or absence of butter.

No missing values were identified in the dataset, ensuring a complete analysis. The technical details of the data indicate it is suitable for factor analysis of variance due to the presence of multiple categorical variables and a quantitative variable.

# 3. Methods

This section provides detailed description of the statistical methods used for analyzing the ant-sandwich attraction data.

## 3.1 Descriptive Statistics

Descriptive statistics are used to summarize and describe the main features of a collection of data. They provide a concise overview of the data, helping to understand its central tendency and dispersion.

**Mean:** The sum of all observations divided by the number of observations. It represents the central value of a set of numbers.

$$\text{Mean} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where x_i are the individual observations and n is the total number of observations.

**Median:** The median of a set of data values for a single quantitative variable, denoted m, is

- the middle entry if an ordered list of the data values contains an odd number of entries, or
- the average of the middle two values if an ordered list contains an even number of entries.

The median splits the data in half.

**Standard Deviation (s):** The standard deviation for a quantitative variable measures the spread of the data in a sample:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

The standard deviation of a sample is denoted s, and measures how spread out the data are from the sample mean x.

The standard deviation gives a rough estimate of the typical distance of a data value from the mean. The larger the standard deviation, the more variability there is in the data and the more spread out the data are.

Graphical representations such as bar charts are used to visualize the mean antCount for different categories of bread, topping, and butter, providing an intuitive understanding of the central tendencies and potential differences between groups.

## 3.2 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical technique used to analyze the differences among group means in a sample. It partitions the total variability in a dataset into different components, allowing researchers to determine if the means of two or more groups are significantly different from each other.

For this experiment, a three-way ANOVA is appropriate as there are three categorical independent variables (bread, topping, butter) and one quantitative dependent variable (antCount).

The general linear model for a three-way ANOVA is given by:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

where:

$Y_{ijkl}$ is the observation for the $l$-th replicate at the $i$-th level of factor A (bread), $j$-th level of factor B (topping), and $k$-th level of factor C (butter).

$\mu$ is the overall mean.

$\alpha_i$ is the main effect of factor A (bread).

$\beta_j$ is the main effect of factor B (topping).

$\gamma_k$ is the main effect of factor C (butter).

$(\alpha\beta)_{ij}$ is the interaction effect between A and B.

$(\alpha\gamma)_{ik}$ is the interaction effect between A and C.

$(\beta\gamma)_{jk}$ is the interaction effect between B and C.

$(\alpha\beta\gamma)_{ijk}$ is the three-way interaction effect among A, B, and C.

$\varepsilon_{ijkl}$ is the random error term, assumed to be $\varepsilon_{ijkl} \sim \mathcal{N}(0,\sigma^2)$.

**Hypotheses for main effects**:

**Null Hypothesis ($H_0$)**: The null hypothesis is that all three types of sandwiches are equally liked by the ants and attract the same mean number of ants.

If we let $\mu_1$, $\mu_2$, and $\mu_3$ represent the mean number of ants on the respective sandwich types, we have

$$H_0: \mu_1 = \mu_2 = \mu_3$$

There is no significant difference in mean ant count across different bread types.

**Alternative Hypothesis ($H_1$)**: The alternative hypothesis is that the means are not all the same. This doesn't imply that all three types of sandwiches have different means, just that the mean for at least one of the fillings is different from the mean for another filling. To state this we use

$$H_a: \text{At least one } \mu_i \neq \mu_j$$

Similar hypotheses are formulated for Topping, Butter, and all interaction terms.

## 3.3. Tukey's HSD Post-hoc Test

When ANOVA indicates a significant main effect for a factor with more than two levels, it only tells us that at least one group mean is different, but not which specific pairs of means are different. Tukey's HSD test is a post-hoc test used to perform all pairwise comparisons among group means while controlling the family-wise error rate (FWER). This means it maintains a specified alpha level (e.g., 0.05) for all comparisons, reducing the chance of Type I errors (false positives) that would occur if multiple t-tests were performed without adjustment (Miller Jr, 1981).

Tukey's HSD calculates a q-statistic for each pair of means and compares it to a critical value. If the calculated q-statistic exceeds the critical value, the difference between the pair of means is considered statistically significant.

# 4. Evaluation

This section presents the results of the descriptive analysis and the statistical hypothesis tests, aimed at answering the research questions.

## 4.1 Descriptive Analysis

The descriptive analysis provides an initial insight into the distribution of antCount across different categories of bread, topping, and butter.

| | antCount | bread | topping | butter |
|---|---|---|---|---|
| **0** | 34 | Rye | Ham and gherkins | no |
| **1** | 47 | Multi Grain | Peanut butter | yes |
| **2** | 67 | White | Ham and gherkins | yes |
| **3** | 63 | Multi Grain | Ham and gherkins | yes |
| **4** | 65 | Rye | Ham and gherkins | no |

Fig.1. Few rows of data

**Overall Ant Count:** The antCount variable ranged from 18 to 76 ants, with an overall mean of 43.50 ants and a standard deviation of 15.15, indicating moderate variability in the number of attracted ants.

| Factor | Category | Mean | Median | Std. Dev. |
|---|---|---|---|---|
| Bread | Whole Grain | 44.5 | 48 | 14.61 |
| | Rye | 44.25 | 42 | 13.4 |
| | Multi Grain | 43 | 40 | 18.23 |
| | White | 42.25 | 44.5 | 15.86 |
| Topping | Ham and gherkins | 55.5 | 58.5 | 12.06 |
| | Peanut butter | 40.38 | 44.5 | 14.18 |
| | Yeast spread | 34.63 | 34.5 | 11.16 |
| Butter | yes | 48.88 | 49 | 14.51 |
| | no | 38.13 | 37 | 14.07 |

Table.1. Descriptive Statistics by Factor

**Bread Type**: The mean ant counts across different bread types are quite similar, ranging from 42.25 (White bread) to 44.50 (Whole Grain). This suggests that bread type might not be a major influencing factor on its own.

**Topping**: There appear to be notable differences in mean ant counts based on topping. "Ham and gherkins" attracted the highest mean number of ants (55.50), followed by "Peanut butter" (40.38), and then "Yeast spread" (34.63). This initial observation suggests that topping might have a significant influence.

**Butter**: Sandwiches "With Butter" attracted a higher mean number of ants (48.88) compared to "Without Butter" (38.13), indicating a potential influence of butter.

The following bar charts visualize the mean ant counts for each factor, with 95% confidence intervals.

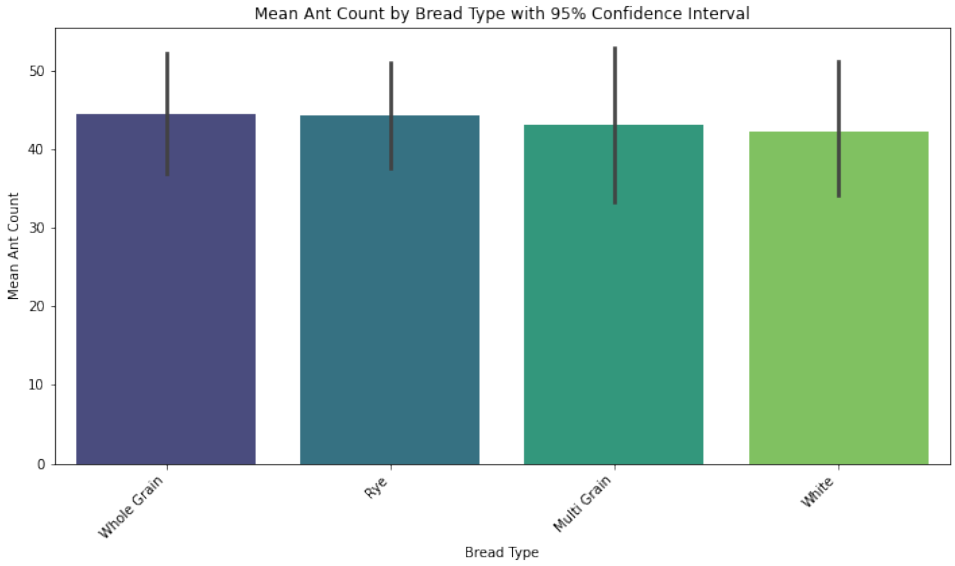


Fig.2. Mean Ant Count by Bread Type with 95% Confidence Interval

As seen in Figure 2, the mean ant counts for different bread types are very close, and their confidence intervals largely overlap, reinforcing the idea of minimal individual impact.

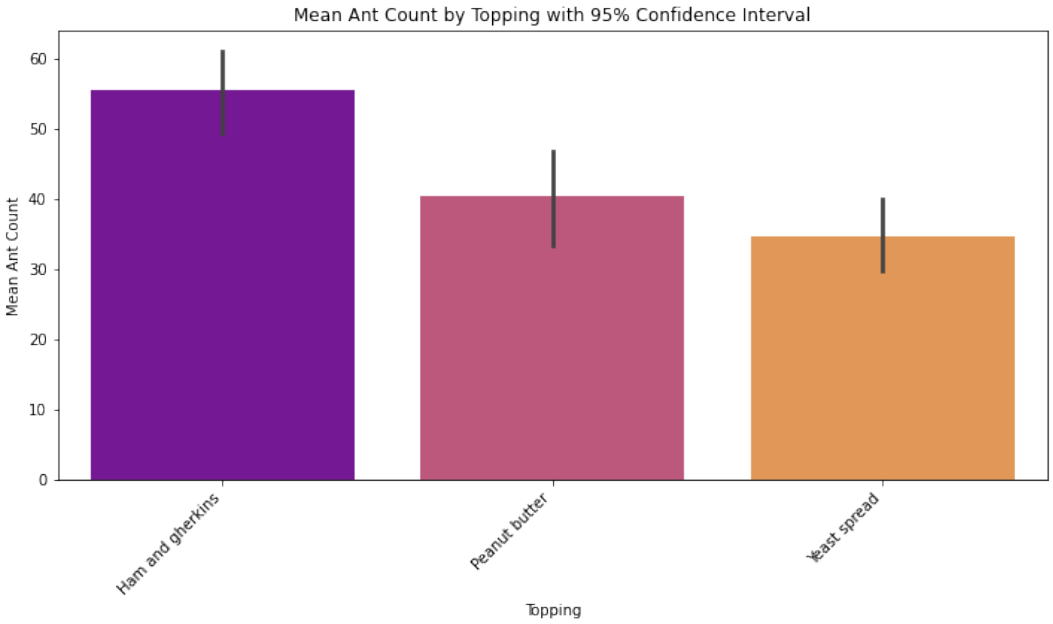


Fig.3. Mean Ant Count by Topping with 95% Confidence Interval

Figure 3 clearly shows distinct differences in mean ant counts for toppings. "Ham and gherkins" appears to attract the most ants, with a considerably higher mean than "Peanut butter" and "Yeast spread."

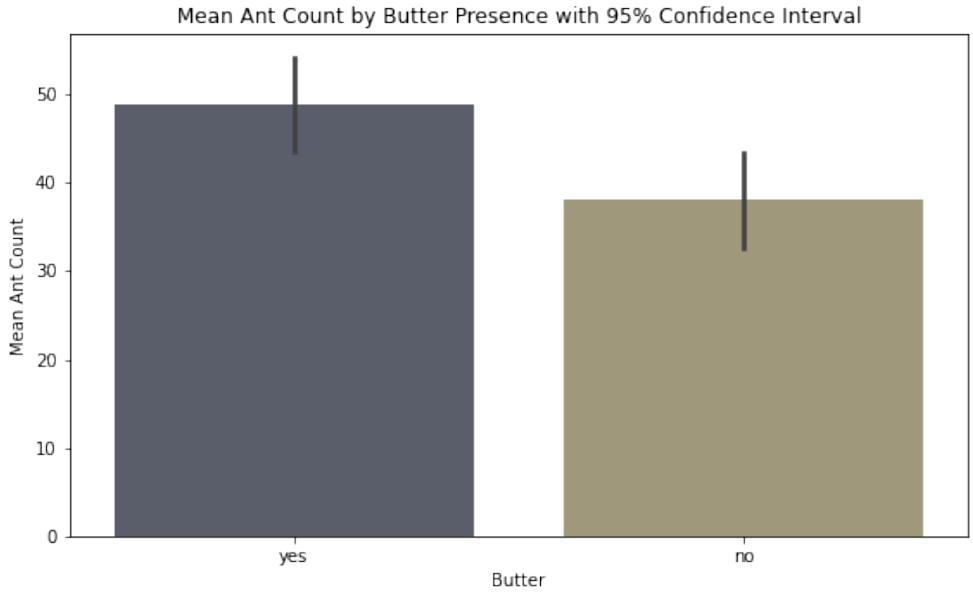


Fig.4. Mean Ant Count by Butter Presence with 95% Confidence Interval

Figure 4 illustrates that sandwiches with butter consistently attracted more ants on average than those without butter, with non-overlapping confidence intervals suggesting a real difference.

## 4.2 Hypothesis Testing Results

To formally assess the significance of the observed differences and potential interactions, a three-way ANOVA was performed.

**ANOVA Results:**

The ANOVA table provides F-statistics and p-values for each main effect and interaction term. The p-value, PR(F), indicates the probability of observing such differences if the null hypothesis were true. A p-value less than the significance level (typically alpha=0.05) leads to the rejection of the null hypothesis.

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| **C(bread)** | 40.50 | 3.0 | 0.142043 | 0.933761 |
| **C(topping)** | 3720.50 | 2.0 | 19.572994 | 0.000009 |
| **C(butter)** | 1386.75 | 1.0 | 14.590969 | 0.000830 |
| **C(bread):C(topping)** | 577.00 | 6.0 | 1.011837 | 0.440987 |
| **C(bread):C(butter)** | 378.75 | 3.0 | 1.328365 | 0.288413 |
| **C(topping):C(butter)** | 56.00 | 2.0 | 0.294608 | 0.747479 |
| **C(bread):C(topping):C(butter)** | 2345.50 | 6.0 | 4.113108 | 0.005597 |
| **Residual** | 2281.00 | 24.0 | NaN | NaN |

Fig.5. ANOVA Table

**Bread Type (C(bread)):** The p-value is 0.933761, which is much greater than 0.05. This indicates that bread type does not have a statistically significant main influence on the number of attracted ants.

**Topping (C(topping)):** The p-value is 0.000009, which is less than 0.05. This indicates that topping has a highly statistically significant main influence on the number of attracted ants.

**Butter (C(butter)):** The p-value is 0.000830, which is less than 0.05. This indicates that the presence of butter has a statistically significant main influence on the number of attracted ants.

**Two-way Interactions:** The p-values for C(bread):C(topping), C(bread):C(butter), and C(topping):C(butter) are all greater than 0.05. This suggests that there are no significant two-way interaction effects between these factors. For example, the effect of bread type on ant attraction does not significantly depend on the topping used, and vice-versa for the other pairs.

**Three-way Interaction:** (C(bread):C(topping):C(butter)): The p-value is 0.005597, which is less than 0.05. This indicates a statistically significant three-way interaction between bread type, topping, and butter. This means that the effect of one factor (e.g., bread type) on ant attraction depends on the levels of the other two factors (topping and butter) in combination.

**Post-hoc Tests (Tukey's HSD):**

Tukey's HSD was performed for bread and topping only. It has not been performed for butter, as it has only two levels (yes or no). Whereas, bread has 4 levels and toppings have 3 levels. When an ANOVA shows a significant main effect for a factor that has only two levels (like 'butter' in your case), you don't need a post-hoc test like Tukey's HSD. The significance itself from the ANOVA for that factor directly tells you that the two levels are significantly different from each other.

Tukey's HSD Post-hoc Test for Bread Type:
Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---|---|---|---|---|---|---|
| Multi Grain | Rye | 1.25 | 0.9 | -15.7854 | 18.2854 | False |
| Multi Grain | White | -0.75 | 0.9 | -17.7854 | 16.2854 | False |
| Multi Grain | Whole Grain | 1.5 | 0.9 | -15.5354 | 18.5354 | False |
| Rye | White | -2.0 | 0.9 | -19.0354 | 15.0354 | False |
| Rye | Whole Grain | 0.25 | 0.9 | -16.7854 | 17.2854 | False |
| White | Whole Grain | 2.25 | 0.9 | -14.7854 | 19.2854 | False |

Table 2: Tukey's HSD Post-hoc Test for Bread Type

As expected from the ANOVA results, none of the pairwise comparisons for bread types are statistically significant (all p-adj > 0.05), confirming that there is no significant difference in ant attraction among the different bread types.

Tukey's HSD Post-hoc Test for Topping:
Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---|---|---|---|---|---|---|
| Ham and gherkins | Peanut butter | -15.125 | 0.0038 | -25.8618 | -4.3882 | True |
| Ham and gherkins | Yeast spread | -20.875 | 0.001 | -31.6118 | -10.1382 | True |
| Peanut butter | Yeast spread | -5.75 | 0.4053 | -16.4868 | 4.9868 | False |

Table 3: Tukey's HSD Post-hoc Test for Topping

The Tukey's HSD test for topping shows the following significant differences:

**"Ham and gherkins" vs. "Peanut butter":** Significant difference (p-adj = 0.0038). "Ham and gherkins" attracted significantly more ants than "Peanut butter" (mean difference of -15.125, meaning "Peanut butter" is lower than "Ham and gherkins").

**"Ham and gherkins" vs. "Yeast spread":** Highly significant difference (p-adj = 0.0001). "Ham and gherkins" attracted significantly more ants than "Yeast spread" (mean difference of -20.875).

**"Peanut butter" vs. "Yeast spread":** No significant difference (p-adj = 0.4037).

Based on these results, "Ham and gherkins" topping attracts the most ants, significantly more than both "Peanut butter" and "Yeast spread". There is no significant difference between "Peanut butter" and "Yeast spread" in terms of ant attraction.

Regarding the question "Which type of bread attracts the most ants?", based on the descriptive statistics, "Whole Grain" had the highest mean ant count, but the ANOVA and Tukey's HSD results show that no specific type of bread significantly attracts more ants than others.

# 5. Summary

This report analyzed an experiment investigating the influence of bread type, topping, and butter on ant attraction to sandwiches, aiming to identify significant factors and the most attractive ingredients.

The descriptive analysis indicated that "Ham and gherkins" topping and the presence of butter were associated with higher ant counts, while bread types showed less variability.

Statistical hypothesis testing via a three-way ANOVA confirmed these observations:

Topping and the presence of butter both had a statistically significant main influence on ant attraction ($p < 0.001$ for both).

Bread type did not show a statistically significant main influence ($p = 0.934$).

A significant three-way interaction between bread type, topping, and butter ($p = 0.006$) was observed, indicating a complex interplay among these factors.

Tukey's HSD post-hoc test for toppings revealed that "Ham and gherkins" significantly attracted more ants than both "Peanut butter" and "Yeast spread." No significant differences were found between "Peanut butter" and "Yeast spread," nor among any of the bread types.

In conclusion, sandwich topping and butter presence are significant drivers of ant attraction, with "Ham and gherkins" topping and sandwiches with butter generally attracting more ants. Bread type, on its own, does not appear to be a key factor. The significant three-way interaction suggests that the combined effect of ingredients is more complex than individual contributions.

These findings suggest that avoiding "Ham and gherkins" and butter may help deter ants. Future research could explore specific chemical attractants, expand ingredient variations, and study ant behavior in diverse environments to generalize these findings.

# 6. Bibliography


- Freund, J. E., & Perles, B. M. (2006). *Statistics: A First Course* (8th ed.). Pearson Prentice Hall. (Used for descriptive statistics definitions).
- Miller Jr, R. G. (1981). *Simultaneous Statistical Inference* (2nd ed.). Springer-Verlag. (Used for Tukey's HSD definition).
- Montgomery, D. C. (2017). *Design and Analysis of Experiments* (9th ed.). John Wiley & Sons. (Used for ANOVA definitions and assumptions).
- Locks (2021). Statistics: *Unlocking the Power of Data (Used for descriptive statistics definitions and references)*