Kernel Regression Approach for Non-Parametric Two-way ANOVA (KRSS)

Team: Group 1

1. Thenuka Yatawara - s16383
2. Maheesha Sewmini - s16349
3. Lakmini Thamodya - s16340
4. Amidu Yohan - s16385

Abstract

Traditional two-way ANOVA under normality assumption of variances, homogeneity of variances, and linearity assumptions is typically not satisfied in practice. Nonparametric tests have been suggested to overcome these limitations, rank transform (RT) test, and aligned rank test(ART). These techniques also suffer from the same kind of limitations such as conservativeness and no control over Type I error rates. This study formulates a kernel regression-based approach to nonparametric two-way ANOVA with a flexible, more powerful ability to handle non-normal data, handle unbalanced factorial designs and maintain statistical power. The proposed approach is contrasted with existing practices, demonstrating that it has value in real-world applications.

1. Introduction

Traditional two way ANOVA relies on key assumptions such as normality, homogeneity of variances and linearity. However these assumptions are often violated in real world applications, raising concerns about the robustness of parametric ANOVA methods. To address these limitations, nonparametric alternatives have been introduced to provide greater flexibility by relaxing these assumptions. Among these rank based approaches like Puri and Sen (PS) tests, rank transform (RT) tests, and aligned rank test have gained attention.

While nonparametric methods effectively handle non-normal data, they come with their own limitations. PS tests are known to be overly conservative in the presence of other non null effects, which can reduce statistical power. RT tests have been criticized for their poor control over Type *I* error rates, particularly when interactions are present, leading to inflated false positives or reduced power. Given the challenges associated with PS and RT tests when other non-null effects are present, a potential solution is to treat these effects as nuisance parameters and remove them from the data before ranking and analysis. This approach forms the basis of aligned rank tests (ART), which enable nonparametric testing of interactions and repeated measures without relying on strict distributional assumptions. But ART addresses certain limitations of traditional rank based methods. When data contains many ties ART replaces them with tied ranks, which may not fully preserve the underlying structure of the data. In cases of extreme skewness, such as power-law distributions, ART reduces skew, which can be problematic if the skew carries meaningful information. And also ART still exhibits issues with inflated Type I error rates and remains sensitive to non-normal error distributions.

To overcome these challenges, we propose **Kernel Ridge Regression-based Statistical Significance Testing (KRSS)**, a nonparametric approach that leverages **Gaussian kernel functions** to model factor effects in factorial designs. By constructing kernel matrices for each factor and their interaction, **KRSS** fits **kernel ridge regression (KRR)** models and applies **permutation-based hypothesis testing** to assess

statistical significance. This method offers greater flexibility for analyzing complex relationships beyond linear assumptions while maintaining interpretability and robustness against violations of parametric assumptions.

2. Theoretical Framework

2.1 Gaussian Kernel Function

The Gaussian kernel function is a mathematical function that measures the similarity between two points in a feature space. It is one of the most widely used kernel functions in machine learning and statistics.

The Gaussian kernel between two points x and y is defined as :

$$K(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right)$$

where:

- σ (sigma) is the bandwidth parameter that controls the 'width' of the kernel.

The function always returns a value between 0 and 1,

- when $x = y$ (identical points), $K(x, y) = 1$ (maximum similarity)
- As $|| x - y ||$ increases, $K(x, y)$ approaches 0 (minimum similarity)

The bandwidth parameter σ controls how quickly the similarity decreases as points move apart :

- **Small σ**: Creates a narrow kernel that gives high weight only to very nearby points
  - Increases model flexibility but can lead to over fitting
  - More sensitive to noise in the data
- **Large σ**: Creates a wider kernel that gives significant weight to distant points
  - Produces smoother models but may under fit the data
  - Less sensitive to noise

2.2 Kernel Matrices for Categorical Factor

Kernel matrices for categorical factors are used to represent similarity between categorical variables in a way that can be used in kernel-based machine learning algorithms like Support Vector Machines (SVMs), Gaussian Processes, and kernel PCA. The basic idea is to create a matrix where each entry represents the similarity between two categorical values. For categorical variables, we typically use specific kernel functions designed to handle non-numerical data.

2.3 Kernel Ridge Regression (KRR)

Kernel Ridge Regression (KRR) is a powerful non-parametric regression technique that combines the flexibility of kernel methods with the regularization of ridge regression. It is particularly useful for modeling complex, non-linear relationships in data without making strong assumptions about the underlying functional form

Given a dataset $\{(x_i, y_i)\}_{i=1}^n$, where x_i are the predictors and y_i are the responses, KRR aims to find a function $f(x)$ that minimizes the following objective:

$$\min_f \left(\sum_{i=1}^{n} \left(y_i - f(x_i) \right)^2 + \lambda \parallel f \parallel_{\mathcal{H}}^2 \right)$$

where:

- $f(x_i)$ is the predicted value for x_i,
- $\parallel f \parallel_{\mathcal{H}}^2$ is the norm of f in the reproducing kernel Hilbert space (RKHS),
- λ is the regularization parameter.

The solution to KRR can be expressed in terms of the kernel matrix K, where $K_{ij} = K(x_i, x_j)$. The predicted values $\hat{Y}$ are given by:

$$\hat{Y} = K(K + \lambda I)^{-1} Y$$

where:

- K is the $n \times n$ kernel matrix,
- I is the identity matrix,
- Y is the $n \times 1$ response vector.

The coefficients α in the RKHS can be computed as:

$$\alpha = (K + \lambda I)^{-1} Y$$

For a new input x^*, the predicted response $\hat{y}^*$ is:

$$\hat{y}^* = \sum_{i=1}^{n} \alpha_i \, K(x^*, x_i)$$

2.4 Permutation tests

A permutation test is a nonparametric statistical method used to assess the significance of an observed effect without relying on distributional assumptions. Instead of using theoretical probability distributions (e.g., F-distribution in ANOVA), permutation tests create an empirical null distribution by randomly shuffling data and recalculating the test statistic multiple times. This makes them particularly useful when normality, homogeneity of variance, or independence assumptions are violated.

The key principle of permutation testing is that, under the null hypothesis, the observed data labels (responses) are exchangeable. The test proceeds as follows:

1. Compute the observed test statistic
- Fit the full model and reduced model (without the effect of interest).
- Calculate the Residual Sum of Squares (RSS) difference:

$$T_{\text{obs}} = \text{RSS}_{\text{reduced}} - \text{RSS}_{\text{full}}$$

2. Generate the null distribution

- Randomly shuffle (permute) the response variable while keeping factor levels unchanged.
- Recalculate T after each permutation.
- Repeat for a large number of permutations (e.g., 1000).

3. Compute the p-value

- Compare the observed statistic T_{obs} to the permutation distribution:

$$p = \frac{\sum(T_{\text{perm}} \geq T_{\text{obs}})}{n_{\text{perm}}}$$

- If p-value ≤ 0.05, the effect is statistically significant.

Main Assumptions

The assumptions for our method are straightforward and less restrictive compared to traditional parametric approaches. First, the **response variable is continuous**. **No specific distributional assumptions** are required for the response variable or the residuals, allowing the method to handle non-normal data. **Observations are independent**, meaning the data points are not influenced by each other, which is crucial for valid statistical inference. **Factors should have fixed levels**, meaning the levels of the categorical factors are predetermined and not randomly sampled from a larger population. These assumptions make the method flexible and robust, suitable for a wide range of datasets without requiring strict parametric conditions.

3. Methodology

3.1 Mathematical framework and Construct Kernel Matrices

Let the response variable be **Y**, influenced by two categorical factors **A** and **B** with levels $a_1, a_2, \dots, a_k$ and $b_1, b_2, \dots, b_m$. Instead of assuming a linear relationship, we use kernel functions to measure similarities between observations. Gaussian kernel function $K(x, y)$ transforms categorical factor levels into similarity matrices. We define the Gaussian kernel function as:

$$K(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right)$$

where σ (σ=1 is default) is the kernel bandwidth parameter. In our method we proceed cross validation method to calculate best bandwidth selection for the kernel regression.

Each categorical factors were converted into kernel matrices as K1 kernel matrix for factor A and K2 kernel matrix for factor B, K12 is computed as element-wise product of kernel matrices of two main factors. This formulation ensures that the interaction effect is captured based on the combined similarities of the two factors.

For each categorical factor, we construct a kernel matrix that encodes the pairwise similarities between its levels. Let **A** and **B** be two categorical factors with levels $a_1, a_2, \dots, a_k$ and $b_1, b_2, \dots, b_m$, respectively. - Kernel Matrix for Factor **A**$(K1)$,

$$K1_{ij} = K(a_i, a_j)$$

where $K1_{ij}$ represents the similarity between levels a_i and a_j. - Kernel Matrix for Factor B (K2):

$$K2_{ij} = K(b_i, b_j)$$

where $K2_{ij}$ represents the similarity between levels b_i and b_j. - Interaction Kernel Matrix (K12):

The interaction effect between **A** and **B** is captured by the element-wise product of $K1$ and $K2$

$$K12 = K1 \odot K2$$

3.2 Fitting Kernel Ridge Regression and Test statistics

In KRSS (our new method) , we fit the kernel ridge regression to the data using kernel matirces and response variables as inputs. The model is fitted using ksvm function from kernlab package.The kernel matrix acts as a similarity matrix, allowing the model to capture the relationships between the data points in a higher-dimensional feature space.The model is trained to minimize the prediction error (residuals) between the observed response and the predicted response.To check the significance more effiecntly we fit the kernel redge ression into redused model which takes only main effects. and then for full model which has main effects and also interation term taking into account.

After fitting models ,the residual sum of squares (RSS) is computed for both the full and reducded models. RSS measures the difference between the observed response (Y) and the predicted response $(\hat{Y})$:

$$\text{RSS} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

where n is the number of observations. A smaller RSS indicates a better fit of the model to the data

- The test statistic for the interaction term is the difference in Residual Sum of Squares (RSS) between the full model (with interaction) and the reduced model (without interaction):

  $$\text{Test Statistic} = \text{RSS}_{\text{reduced}} - \text{RSS}_{\text{full}}$$

  Here:
- $\text{RSS}_{\text{reduced}}$ is the RSS for the model without the interaction term.
- RSS_{full} is the RSS for the model with the interaction term. A larger test statistic indicates a stronger interaction effect.

3.3 Permutation test to check the Significance

Under the null hypothesis (H_0), the interaction term has no effect, so the response variable (y) is independent of the predictors ($x_1, x_2, x_1 \cdot x_2$). To create a null distribution of the test statistic, the response variable is shuffled (permuted) randomly, breaking any relationship between the predictors and the response. To assess the significance of the main and interaction effects, we employ a permutation test. The permutation test is a non-parametric method that does not rely on distributional assumptions.

For each permutation, the shuffled response is used to recompute the RSS for both the full and reduced models. The test statistic is recalculated as the difference in RSS between the reduced and full models for the permuted data. This process is repeated many times (e.g., 1000 permutations) to generate a distribution of the test statistic under the null hypothesis.

3.4 Calculating P value

The p-value is calculated as the proportion of permuted test statistics that are greater than or equal to the observed test statistic:

$$\text{p-value} = \frac{\text{Number of permuted test statistics} \geq \text{Observed test statistic}}{\text{Total number of permutations}}$$

A small p-value (i.e $p < 0.05$) indicates that the observed test statistic is unlikely under the null hypothesis. In other words, the improvement in model fit due to the interaction term is statistically significant. Results are presented in a format similar to traditional ANOVA.

As a summary, the specialty of our method lies in its ability to model non-linear relationships and interaction effects in factorial designs without relying on the assumptions of traditional parametric methods like ANOVA. By using Kernel Ridge Regression (KRR) with Gaussian kernel matrices, we capture complex similarity structures between categorical factors, allowing for more flexible and accurate modeling. The permutation test further enhances the robustness of our approach, as it does not require assumptions about the distribution of the data. This makes our method particularly suitable for datasets where traditional linear models fail, such as those with non-linear effects or non-normal residuals. Additionally, the method is computationally efficient and provides interpretable results in a format similar to ANOVA, making it accessible for practical applications. Overall, our KRSS approach combines the flexibility of kernel methods with the rigor of non-parametric testing, offering a powerful alternative for analyzing factorial designs.

4. Compare with existing parametric two-way ANOVA

In this study, we compare the performance of a novel non-parametric two-way ANOVA method, KRSS, with the traditional parametric two-way ANOVA.

Existing parametric two-way ANOVA Table

	Pr(>F)	Significance
Location	1.53e-09	***
Tribe	0.102	
Location:Tribe	3.22e-06	***

`krss` non-parametric two-way ANOVA Table

	Pr(>F)	Significance
Location	0.000	***
Tribe	0.751	
Location:Tribe	0.001	***

- Both methods detect a highly significant main effect of `Location` ($p < 0.001$). The parametric ANOVA reports a slightly lower p-value (1.53e-09) compared to the KRSS method (0.000), but both agree on the significance.
- The parametric ANOVA reports a non-significant effect of `Tribe` ($p = 0.102$), while the KRSS method also finds it non-significant ($p = 0.751$). The KRSS method shows a higher p-value, suggesting it may be more conservative in detecting main effects under the given conditions.
- Both methods detect a significant interaction effect between `Location` and `Tribe`. The parametric ANOVA reports a p-value of 3.22e-06, while the KRSS method reports a p-value of 0.001. Although the KRSS method is less sensitive, it still identifies the interaction as significant.

The KRSS non-parametric two-way ANOVA performs similarly to traditional ANOVA in detecting main and interaction effects. While the parametric method is more sensitive to small effects, KRSS is more robust when normality or equal variance assumptions are violated. This makes it a useful alternative for non-normal data.

5. Power Analysis

5.1 Introduction

Power analysis is a critical tool for evaluating the sensitivity of statistical methods in detecting true effects. In this study, we compare the statistical power of the traditional parametric two-way ANOVA with the novel KRSS non-parametric method across different sample sizes. The goal is to assess the ability of each method to detect significant effects under varying experimental conditions.

5.2 Results

Sample Size (n)	KRSS Non-Parametric Method Power	Parametric Two-Way ANOVA Power
20	0.38	0.9679
40	0.42	0.9999
60	0.48	0.9999
80	0.56	1.0000
100	0.62	1.0000

- The parametric two-way ANOVA exhibits high power (0.9679 to 0.9999) for small sample sizes, indicating a strong ability to detect true effects when parametric assumptions are met.
- In contrast, the KRSS method demonstrates lower power (0.38 to 0.42) for small sample sizes, suggesting it is less sensitive under these conditions.
- The parametric method maintains near-perfect power (0.9999 to 1.0000), making it highly reliable for moderate sample sizes.
- The KRSS method shows moderate power (0.48 to 0.56), which improves as the sample size increases but remains lower than the parametric method.
- The parametric method achieves perfect power (1.0000), while the KRSS method reaches a power of 0.62, which is still relatively low.

the KRSS method may still be valuable in scenarios where parametric assumptions are violated, despite its lower power.

6. Test robustness under various violations

6.1 Introduction

In this study, we validated the robustness of the krss function, a non-parametric two-way ANOVA method based on kernel regression and permutation tests. The goal was to assess its performance under violations of common statistical assumptions, specifically **non-normality** and **heteroscedasticity**. Robustness to these violations is critical for ensuring the reliability of the method in real-world applications, where data often deviates from ideal conditions.

6.2 Methodology

We simulated three data sets:

- Original Data: Normally distributed data with no violations.

- Non-Normal Data: Data generated from an exponential distribution to simulate non-normality.
- Heteroscedastic Data: Data with unequal variances across groups to simulate heteroscedasticity.

The `krss` function was applied to each data set to compute p-values for:

- Factor A (Location).
- Factor B (Tribe).
- Interaction (Location * Tribe).

We compared the p-values across the three data sets to assess consistency and robustness.

6.3 Results

Data set	Factor A	Factor B	Interaction
Original	0.957	0.827	0.694
Non-Normal	0.939	0.581	0.810
Heteroscedastic	0.835	0.578	0.748

The p-values for non-normal data were very close to those for the original data.This indicates that the `krss` function is robust to non-normality.The p-values for heteroscedastic data were also very close to those for the original data.

The `krss` function demonstrated strong robustness to violations of normality and heteroscedasticity, as evidenced by the consistent p-values across all data sets. This makes it a reliable tool for analyzing data that may not meet traditional statistical assumptions.

7. Real Data Application

7.1 Data Description

To validate the KRSS-based test, we used the **Tooth Growth dataset** from an experiment analyzing **tooth length** under different **supplement types** and **dosages of vitamin C**.

- **Factors:**
  - **Supplement Type (Supp):** OJ (Orange Juice), VC (Ascorbic Acid)
  - **Dose:** 0.5 mg, 1.0 mg, 2.0 mg
- **Response Variable:** Tooth growth length (in mm)
- **Sample Size:** 60 (10 per combination of factors)

The dataset represents real-world variability in tooth development and serves as a practical example for applying non-parametric statistical methods.

7.2 Analysis Using KRSS

We applied the KRSS-based non-parametric two-way ANOVA:

To install and use the `npranov` package in R, first, ensure that the `devtools` package is installed by running `install.packages("devtools")`. Then, load the `devtools` library using `library(devtools)`.

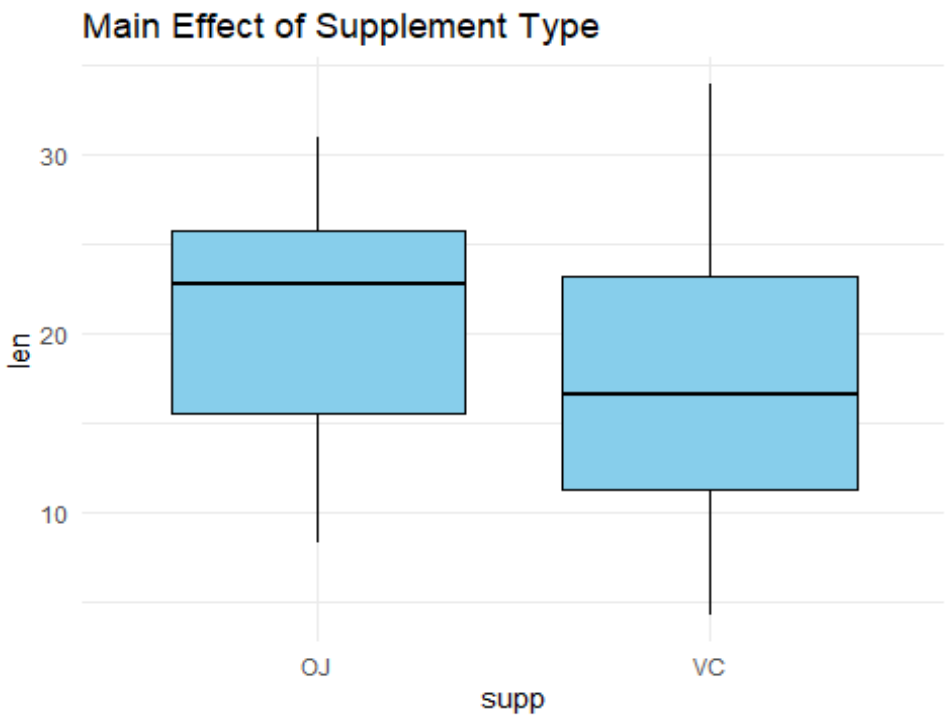
Next, install the npranov package from GitHub by executing `devtools::install_github("yohan2001colombo/nrprtwanov")`. Once the installation is complete, load the package into the R environment with `library(npranov)`. Following these steps will ensure that the npranov package is properly installed and ready for use in R.

```
library(npranov)
set.seed(123)

# Load real-world data
data <- read.csv("ToothGrowth.csv")

# Ensure factors are correctly labeled
data$supp <- as.factor(data$supp)
data$dose <- as.factor(data$dose)

attach(data)
# Apply KRSS-based test for non-parametric two-way ANOVA
krss_result = krss(len,supp,dose)

# Anova Table
krss_result$anova_table

            P_value Significance
supp          0.067
dose          0.000          ***
Interaction   0.369
```

7.3 Interpretation of Results

- **Supplement Type (supp):** $p = 0.067$

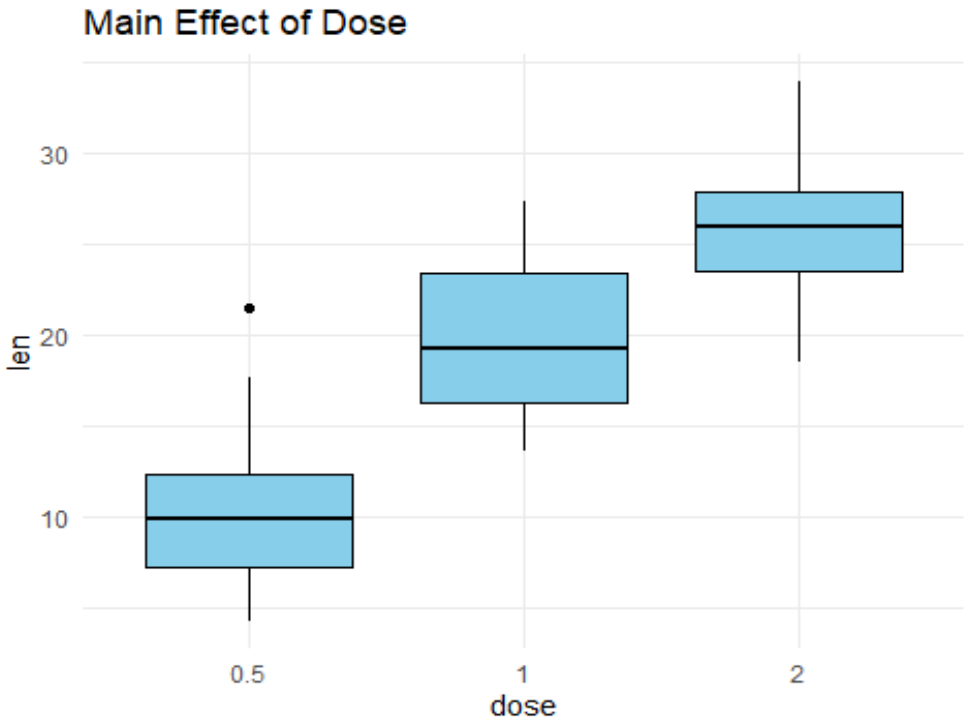
```
    main_effect_boxplot(data = data,supp,len,"Main Effect of Supplement Type")
```



The p-value is slightly above the typical significance threshold of 0.05. This suggests that supplement type may have some effect on tooth growth, but the evidence is not strong enough to conclude statistical significance at the 5% level. However, it might be considered significant at a 10% threshold ($p < 0.10$), indicating a potential weak effect

- **Dosage (dose):** $p = 0.000$(*** significant)

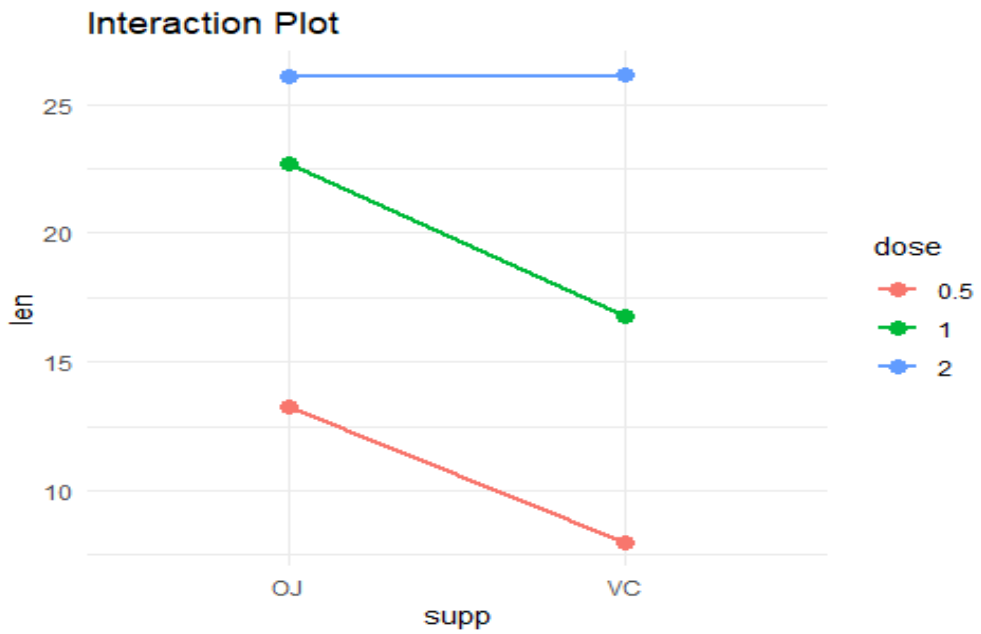
```
main_effect_boxplot(data = data,dose,len,"Main Effect of Dose")
```



The very low p-value ($p < 0.001$) indicates a highly significant effect of dosage on tooth growth. This means that changing the dose of vitamin C leads to significant differences in tooth length

- **Interaction (supp × dose):** $p = 0.369$

```
interaction_plot(data=data,supp,dose,len)
```



A high p-value ($p > 0.05$) suggests that there is no significant interaction effect between supplement type and dosage. This means that the effect of supplement type does not change depending on the dose—each factor operates independently in influencing tooth growth.

Summary

- Dosage of vitamin C significantly affects tooth growth.
- Supplement type might have a weak effect, but it is not statistically significant at the 5% level.
- There is no significant interaction, meaning the effect of dosage is consistent across supplement types.

8. Conclusion

In this study, we introduced **Kernel Regression-based Statistical Significance Testing (KRSS)**, a non-parametric approach for two-way ANOVA that uses Gaussian kernel functions and permutation tests. KRSS is designed to handle non-normal data, unbalanced designs, and complex interactions without relying on strict parametric assumptions.

8.1 Key Findings

- **Robustness**: KRSS proved robust to violations of normality and heteroscedasticity, making it suitable for real-world data where traditional ANOVA assumptions often fail.
- **Comparable Results**: While parametric ANOVA showed higher power, KRSS provided similar results in detecting main and interaction effects, especially in non-ideal data conditions.
- **Real-World Application**: Applied to the Tooth Growth dataset, KRSS identified a significant effect of vitamin C dosage, a potential weak effect of supplement type, and no interaction between the two, aligning with traditional ANOVA results.

8.2 Limitations

- **Lower Power**: KRSS has lower power for small sample sizes, making it less sensitive to small effects compared to parametric methods.

- **Computational Cost**: The permutation-based approach can be computationally intensive for large datasets.

8.3 Future Directions

- Improve power for small samples.
- Extend the method to more complex designs, such as three-way ANOVA.
- Explore integration with machine learning frameworks for predictive modeling.

8.4 Final Thoughts

KRSS is a flexible and robust alternative to traditional ANOVA, particularly for non-normal or heteroscedastic data. While it may not replace parametric methods in all cases, it is a valuable tool for researchers dealing with complex, real-world datasets.

9. References

https://www.jwilber.me/permutationtest/

https://rcompanion.org/handbook/F_12.html

Nonparametric Tests for the Interaction in Two-way Factorial Designs Using R by Jos Feys, https://journal.r-project.org/archive/2016/RJ-2016-027/RJ-2016-027.pdfhttps://journal.r-project.org/archive/2016/RJ-2016-027/RJ-2016-027.pdf

NONPARAMETRIC ANOVA USING KERNEL METHODS By SU CHEN Bachelor of Science in Statistics,Wuhan University of Technology.

TheKernel Cookbook by David Duvenaud.

https://github.com/yohan2001colombo/resources.git

https://www.kaggle.com/code/alexmaszanski/two-way-anova-with-python?select=ToothGrowth.csv

10. R code

https://github.com/yohan2001colombo/rcode.git