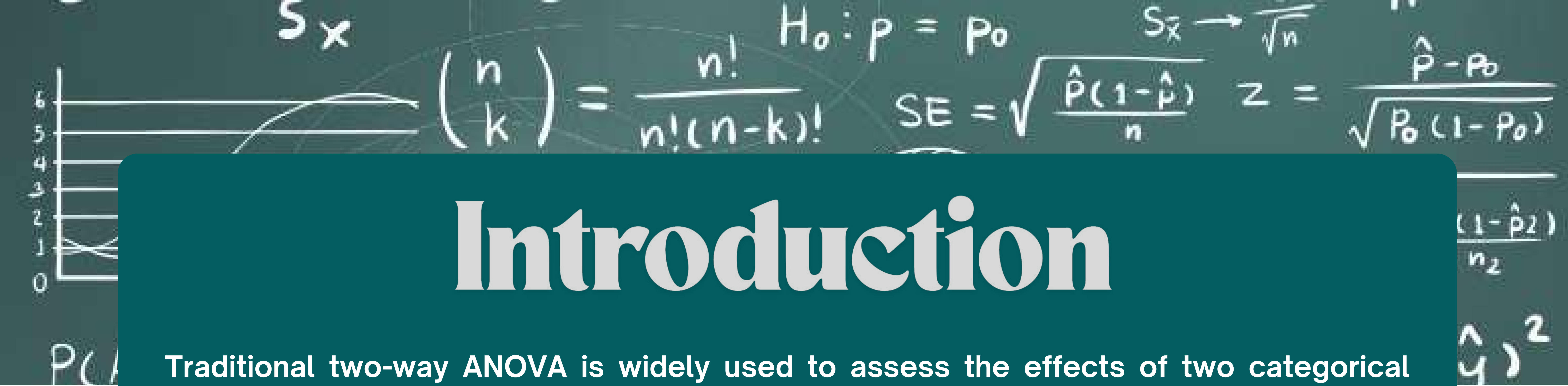
Alternative Method for Non-Parametric Two-way ANOVA

Group 04



Introduction

Traditional two-way ANOVA is widely used to assess the effects of two categorical variables on a continuous outcome, but it relies on assumptions like normality and equal variances. This study develops a non-parametric alternative using a bootstrap-permutation approach that:

- Captures interaction effects,
- Avoids strict parametric assumptions,
- Estimates effect sizes using Eta Squared (η^2)
- Uses resampling for robust inference.

TWO-WAY ANOVA

- In factorial experiments, two-way ANOVA examines the effects of two categorical independent variables on a dependent variable.

- However traditional ANOVA assumes normality, homogeneity of Variance, and independence of observations.

- When these assumptions are violated, non parametric alternatives are preferred.

Theoretical Findings

Aligned Rank Transform (ART) ANOVA

WHAT IS ALIGNED RANK TRANSFORM (ART) ANOVA ?

- Aligned Rank Transform (ART) ANOVA is a non-parametric statistical method that extends the ANOVA framework to handle ordinal or non normal data.
- It aligns and ranks data for each effect, enabling analysis of factorial designs without assuming normality or equal variances.

Permutation Tests

WHAT ARE PERMUTATION TESTS IN TWO-WAY ANOVA?

- A permutation test for two-way ANOVA, often referred to as PERMANOVA (Permutational Multivariate Analysis of Variance), serves as a non-parametric alternative to the traditional two-way ANOVA.

- Unlike standard methods, which assume normality and equal variance, permutation tests rely on randomly shuffling the data labels to test whether the observed group differences are statistically significant.

Novel Methodology

Step 01 : Assumptions

1. Exchangeability of Observations
2. Fixed Factor Structure
3. Sufficient Number of Permutations

Step 02 : Define the hypotheses

- Null Hypothesis (H0): There are no significant differences among the groups.
- Alternative Hypothesis (Ha): At least one group shows a significant difference.

Step 03 : Compute Observed Test Statistic (Eta Squared)

- The test statistic used is Eta Squared (η^2), which is defined as

$$\boxed{\eta^2 = \frac{SS_{\text{Effect}}}{SS_{\text{Total}}}}$$

where: SS_Effect represents the sum of squares for the effect of interest (main effects or interaction). SS_Total is the total sum of squares in the dataset.

Eta Squared (η^2)

Eta Squared measures the proportion of total variance explained by a factor.

Why Use Eta Squared (η^2) ...?

- **Effect Size Measure**
- **Interpretability**

0	→	No effect
0.01	→	Small effect
0.06	→	Medium effect
0.14+	→	Large effect (Cohen, 1988)

- **Consistency Across Bootstrap & Permutation**

Assumptions

- *None required for interpretation.*
- *Suitable for non-parametric methods and bootstrap-based resampling approaches.*

Step 04 : Permutation Testing for Significance

- To determine p-values, permutation testing is conducted by random shuffling of data labels and recalculating η^2 under the null hypothesis.

Permutation Procedure

1. **Shuffle labels**: Factor labels (A and B) are randomly permuted while keeping η^2 fixed.
2. **Recompute** η^2: The permuted dataset's η^2 is calculated.
3. **Repeat for P permutations** (e.g., P = 5000) to build the null distribution.
4. **Compute p-values**: The proportion of times η^2 exceeds the observed η^2 is calculated as:

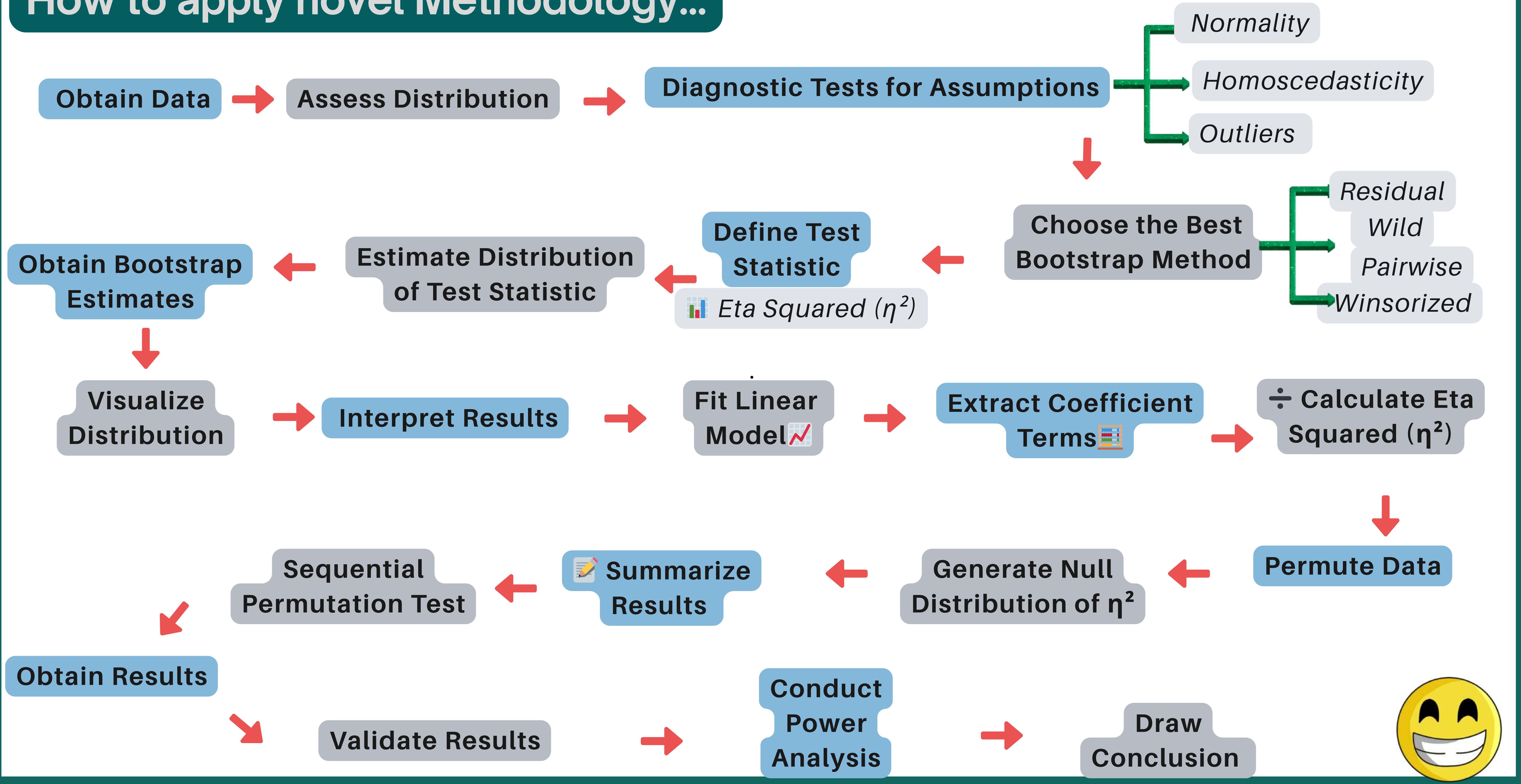
$$p = \frac{\sum_{p=1}^{P} I\left(\eta^2_{\text{perm},p} \geq \eta^2_{\text{obs}}\right)}{P}$$

where $I(.)$ is an **indicator function** that equals 1 if $\eta^2_{perm} >= \eta^2_{obs}$, otherwise 0.
- η^2_{obs} is the observed effect size.

- If the p-value is less than the significance level (e.g., α = 0.05), reject the null hypothesis (H0) and conclude that there is a significant effect.
- If the p-value is greater than or equal to the significance level, fail to reject the null hypothesis (H0), suggesting no strong evidence for a significant effect.

Concepts and Implementation of Novel Methodology

How to apply novel Methodology...



Step 1 : Bootstrapping

WHY WE USE BOOTSTRAPPING.....?

- Estimates the sampling distribution of η^2 without relying on normality.
- Generates confidence intervals and standard errors through resampling.
- Effectively handles :
  - Non-normal data
  - Unequal variances (heteroscedasticity)
  - Violations of ANOVA assumptions

Step 1 : Bootstrapping

1. Choosing the Best Bootstrap Type

→ **Multiple non-parametric bootstrap methods are executed to handle different conditions in the data, selecting the best method based on the results of assumption checks.**

Residual Bootstrap

If normality holds

Resample residuals to create new datasets

Wild Bootstrap

If there is heteroscedasticity

Adjust residuals using random sign flips to maintain heteroscedasticity

Pairwise Bootstrap

If normality is violated

Resample entire rows of data instead of residuals

Winsorized Bootstrap

If there is heteroscedasticity & Outliers exist

Trim extreme residuals to reduce outlier influence

2. Define Test Statistic: Eta Squared

$$\eta^2 = \frac{\mathrm{SS}_{\text{effect}}}{\mathrm{SS}_{\text{total}}}$$

Where :

$\mathrm{SS}_{\text{effect}} = \sum(\hat{Y}_i - \bar{Y})^2$ --the sum of squares for the factor.

$\mathrm{SS}_{\text{total}} = \sum(Y_i - \bar{Y})^2$ --the total sum of squares.

- Eta squared tells us how much of the variation in the data is explained by the factor we're testing.It’s calculated as the sum of squares for the effect, divided by the total sum of squares.

3. Estimate distribution of test statistic

- Now, once the test statistic is defined, we estimate its distribution using the appropriate bootstrap method:

Residual Bootstraping

$$Y_i^* = \hat{Y}_i + e_i^*$$

- $\hat{Y}_i$: fitted value from the original model
- e_i^*: randomly sampled residual (with replacement) from the original residuals $e_i = Y_i - \hat{Y}_i$

Wild Bootstraping

$$Y_i^* = \hat{Y}_i + e_i \cdot w_i$$

- $w_i \in \{-1, +1\}$: randomly sampled weights for each i, with equal probability
- Keeps the variance pattern (heteroscedasticity) while resampling

Pairwise Bootstraping

$$(Y_i^*, A_i^*, B_i^*) \sim \text{sample with replacement from original rows}$$

Winsorized Bootstraping

1. Define bounds:

$$L = \text{Quantile}(e, \alpha), \quad U = \text{Quantile}(e, 1 - \alpha)$$

2. Winsorize residuals:

$$\tilde{e}_i = \min(\max(e_i, L), U)$$

3. Resample from these:

$$Y_i^* = \hat{Y}_i + \tilde{e}_i^*$$

4. Execution of Bootstrap Procedure

- Apply the optimal bootstrap method to estimate η^2 (effect size), producing:
  - Confidence Intervals
  - Standard Error and Mean

5. Visualization of Bootstrap Distribution

- Used ggplot to display the bootstrap distribution of η^2.
- The peak of the curve shows where most values fall, telling us how strong the effect is.

6. Interpretation of Results

Mean η^2 : Avg. variance explained across samples
Confidence Interval : Range of plausible η^2 values
Standard Error : Variability in η^2 estimates
Density Plot : Confirms reliability & stability of results

Step 2 : Permutation Approach

WHY WE USE PERMUTATION.....?

- The purpose of permutation testing is to assess the effect of independent factors A, B and their interactions A*B, on the dependent variable Y, by randomly assigning values and recalculating the effect size; eta squared.

WHAT ARE THE KEY FUNCTIONS OF PERMUTATION APPROACH ?

1. Coefficient Extraction
2. Eta squared calculation
3. Data permutation
4. Permutation Execution
5. Summarization of Results
6. Sequential Permutation Testing

1. Coefficient Extraction

- Using this function, the coefficient name for a given term from the fitted linear model is extracted. This is later used to find the exact terms that are needed to check significance.

2. Eta squared calculation

- The proportion of variance explained by factors A, B, and interaction A*B is computed using the ANOVA sum of squares.

$$SS_{\text{Total}} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

- **Main effect of A:**

$$\eta_A^2 = \frac{SS_A}{SS_{\text{Total}}}$$

- **Main effect of B:**

$$\eta_B^2 = \frac{SS_B}{SS_{\text{Total}}}$$

- **Interaction effect (A × B):**

$$\eta_{A:B}^2 = \frac{SS_{A:B}}{SS_{\text{Total}}}$$

3. Data permutation

Here data is randomly shuffled based on the type of permutation.This means we randomly shuffle data depending on what we want to test.This helps us create new versions of the dataset

- "main_A": Permute factor A while other variables are kept constant.
- "main_B ": Permute factor B while other variables are kept constant.
- "interaction" : Permute dependent Y to test interaction.

- Let D be the orginal dataset and D* is the new dataset.

    - $D = \{(A_i, B_i, Y_i)\}_{i=1}^{n}$: Original dataset
    - $\text{type} \in \{\text{"main\_A"}, \text{"main\_B"}, \text{"interaction"}\}$

- Let π(·) be a random permutation function.Depending on the type, generate a new dataset D* as follows:

◆ 1. If `type = "main_A"` :

- Permute factor A:

$$A_i^* = \pi(A_i), \quad B_i^* = B_i, \quad Y_i^* = Y_i$$

◆ 2. If `type = "main_B"` :

- Permute factor B:

$$A_i^* = A_i, \quad B_i^* = \pi(B_i), \quad Y_i^* = Y_i$$

◆ 3. If `type = "interaction"` :

- Permute response Y:

$$A_i^* = A_i, \quad B_i^* = B_i, \quad Y_i^* = \pi(Y_i)$$

$$D^* = \{(A_i^*, B_i^*, Y_i^*)\}_{i=1}^{n}$$

4. Permutation Execution

- The Permutation process is repeated B times (generally 1000) to generate a null distribution of eta squared values for each effect. This is later compared with the observed distribution of eta squared.

5. Summarization of Results

- Here, the mean, standard deviation, and confidence intervals of permuted eta squared values are computed.

6. Sequential Permutation Testing

- The permutation test is performed sequentially, shuffling A, B, and interaction terms, fitting the model and calculating eta squared distributions.

Validation: Power Analysis

- Power Analysis estimates the probability that a statistical test will detect an effect.

$$\text{Estimated Power} = \frac{\text{Number of simulations with } p\text{-value} < \alpha}{\text{Total number of simulations}}$$

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(p_i < \alpha)$$

- We calculate Estimated Power by checking how many simulations give us a significant p-value (less than alpha), divided by the total number of simulations.

A power of 0.80 is conventionally considered adequate, as recommended by Cohen (1988), to balance Type I and Type II error risks in hypothesis testing.

Procedure :

1. **Generate Data**

2. **Run Permutation Test:** Apply a non-parametric test to each dataset using B permutations.

3. **Estimate Power:** Repeat simulations in parallel and calculate power as the proportion of significant results.

Power Analysis for Simulated Data

- 200 observations
- Non-normality detected → Pairwise Bootstrap Method
- Significance results validated by all 3 ANOVA methods

effect size	n per group	no. of simulations	no. of replicates	estimated power for A	estimated power for B	estimated power for A: B
0.1	50	100	500	0.08	0.03	0.06
0.5	40	100	500	0.34	0.35	0.28
0.7	50	100	500	0.66	0.62	0.69
0.8	50	200	500	0.835	0.77	0.745

Validation: Comparison and Testing Robustness

- The results obtained from Sequential Permutations can be compared with the existing methods for validation of results.
- Robustness is tested by applying for data where traditional assumptions are violated.
- Results for Synthetic data with heteroscedasticity and presence of outliers:

Our method

```
                 Term Eta_Squared P_Value Significance
1                   A  0.35003727   0.000          ***
2                   B  0.28570660   0.000          ***
3 A:B (Interaction)  0.03733653   0.009           **
```

Traditional ANOVA

```
Analysis of Variance Table

Response: Y
           Df Sum Sq Mean Sq F value    Pr(>F)    
A           1 366.81  366.81 209.860 < 2.2e-16 ***
B           1 299.39  299.39 171.291 < 2.2e-16 ***
A:B         1  39.13   39.13  22.385 4.264e-06 ***
Residuals 196 342.58    1.75                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ART ANOVA

```
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Anova Table (Type III tests)
Model: No Repeated Measures (lm)
Response: art(Y)

      Df Df.res F value     Pr(>F)    
1 A    1    196 314.988 < 2.22e-16 ***
2 B    1    196 225.081 < 2.22e-16 ***
3 A:B  1    196  41.136 1.0401e-09 ***
---
Signif. codes:   0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Validation: Comparison and Testing Robustness

- ‘Diamonds Dataset’ sourced from Kaggle consisting 53,940 observations
- Response - Diamond Price, Predictors - Diamond Cut, Diamond Color
- Results for data with heteroscedasticity and presence of outliers:

Our method

```
                   Term Eta_Squared P_Value Significance
1                     A 0.012862074       0          ***
2                     B 0.029712105       0          ***
3 A:B (Interaction) 0.001926042       0           **
```

Traditional ANOVA

```
Response: Y
             Df     Sum Sq    Mean Sq  F value    Pr(>F)
A             4 1.1042e+10 2760436340 181.4051 < 2.2e-16 ***
B             6 2.5507e+10 4251174041 279.3706 < 2.2e-16 ***
A:B          24 1.6535e+09   68893962   4.5274 1.001e-12 ***
Residuals 53905 8.2027e+11   15216972
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ART ANOVA

```
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Anova Table (Type III tests)
Model: No Repeated Measures (lm)
Response: art(Y)

      Df Df.res F value     Pr(>F)
1 A    4  53905 234.722 < 2.22e-16 ***
2 B    6  53905 158.273 < 2.22e-16 ***
3 A:B 24  53905  12.527 < 2.22e-16 ***
```

Implementation Challenges

Bootstrap, Permutation and Computational trade-off

1. Computational Cost

- Bootstrapping and Permutations take long especially with large datasets + many resamples.
- Sequential dependencies make parallelization tough. (Efron and Tibshirani, 1994)

2. Power Analysis Function

- Running hundreds of simulations is brutal on time and memory.
- Low power for small or subtle effects.

3. Significance Assessment Choice

- Initially used Coefficient-based test which is difficult to compare across methods.
- Switched to p values for eta-squared.

Results And Conclusions

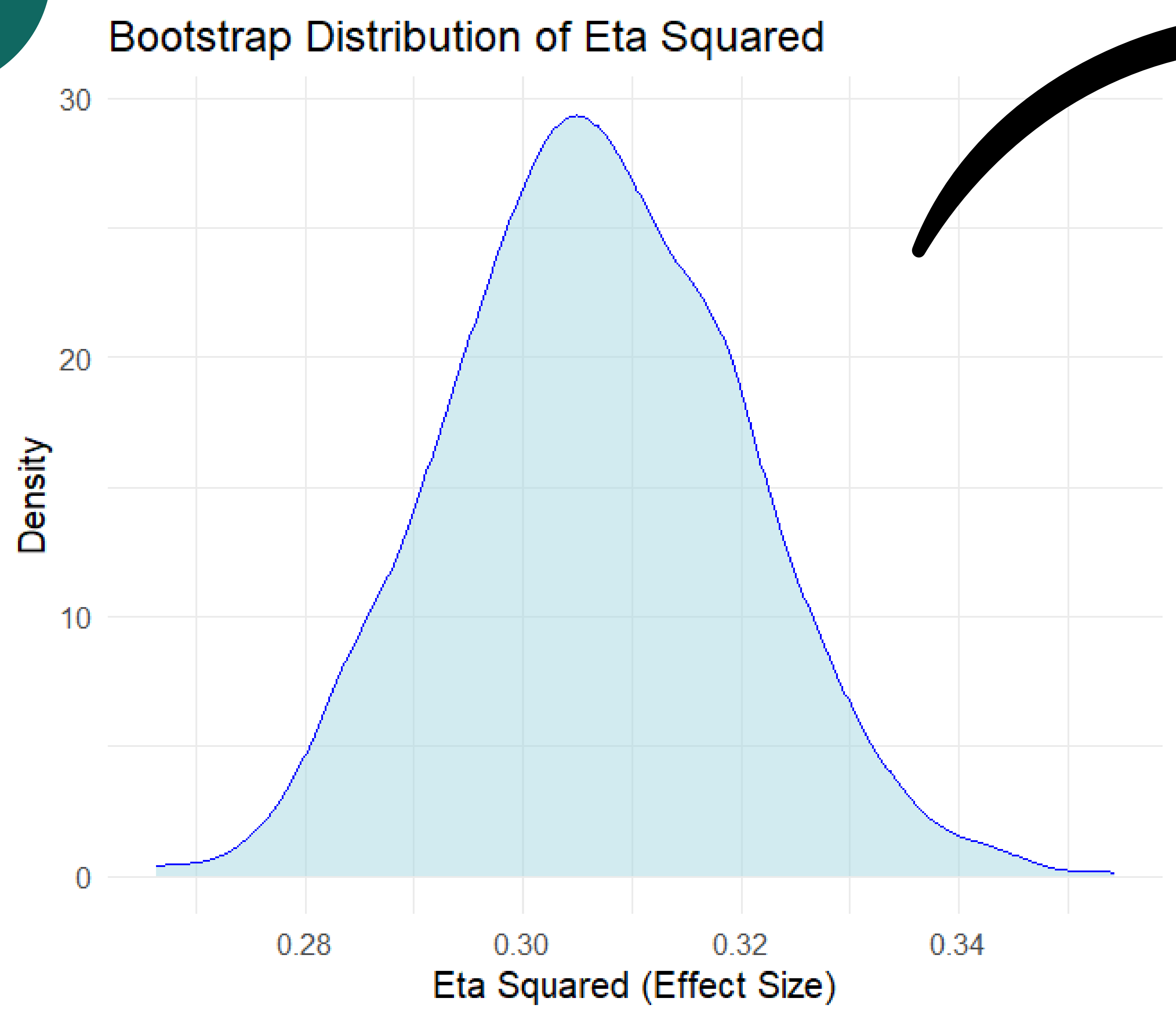
Real Data Application in R

📊 Dataset Overview

- Source: Kaggle (2930 observations)
- Link: https://www.kaggle.com/code/brekhnaa/two-way-anova-and-interactions?select=AmesHousing.csv
- Response Variable: Sale Price
- Factors:
  - Factor A: Season – 4 levels (Spring, Summer, Fall, Winter)
  - Factor B: Heating Quality – 5 levels (Poor, Fair, Average, Good, Excellent)

SEASON	Ex	Fa	Gd	Po	Ta
Autumn	244	14	71	1	147
Spring	448	34	150	0	274
Summer	612	32	200	0	343
Winter	191	12	55	2	100

Bootstrap Approach



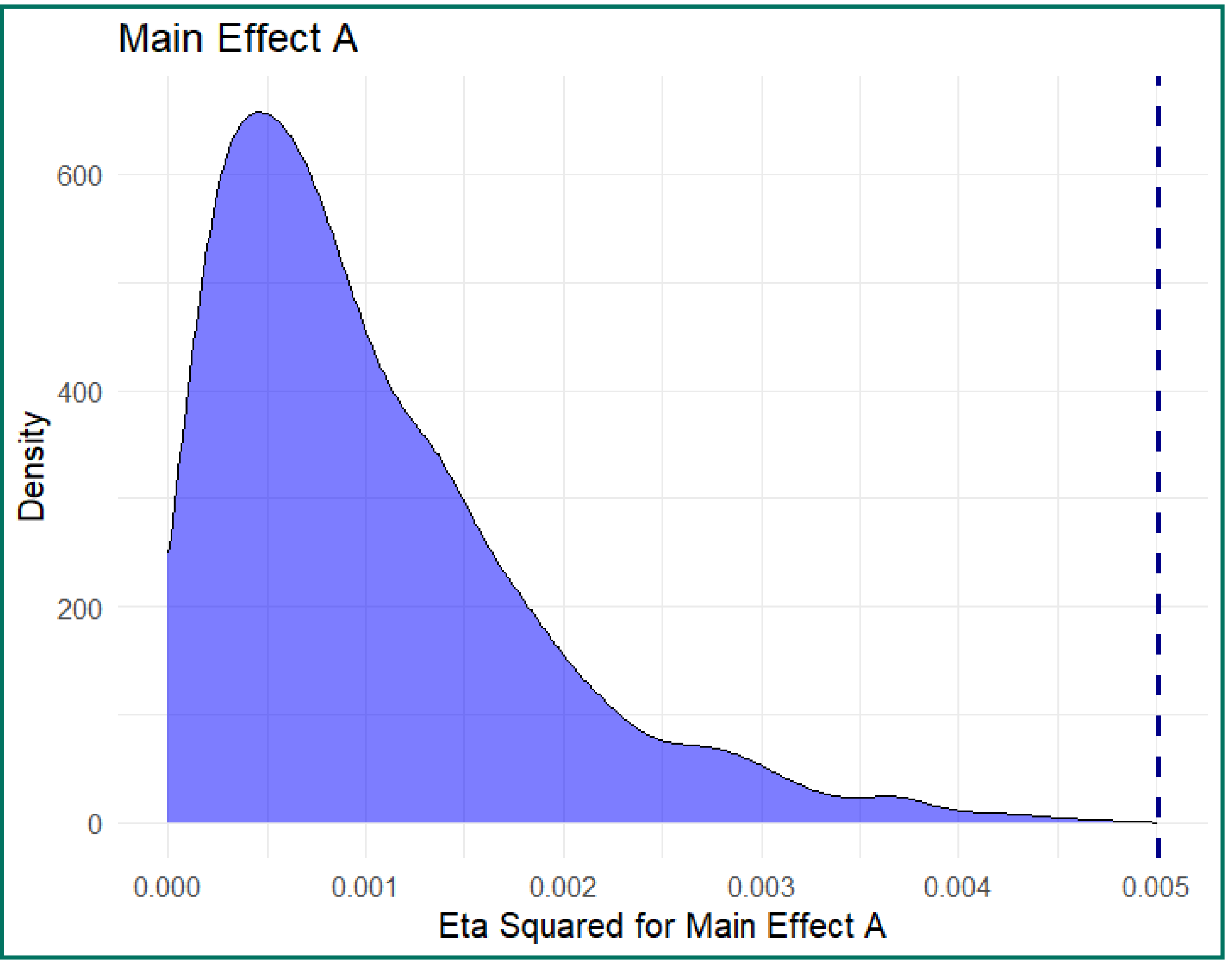
Bootstrap Method : Winsorized
Mean eta squared : 0.3073
Standard deviation eta squared: 0.0131
Lower CI eta squared: 0.2812
Upper CI eta squared: 0.3333

Assumption Violations:

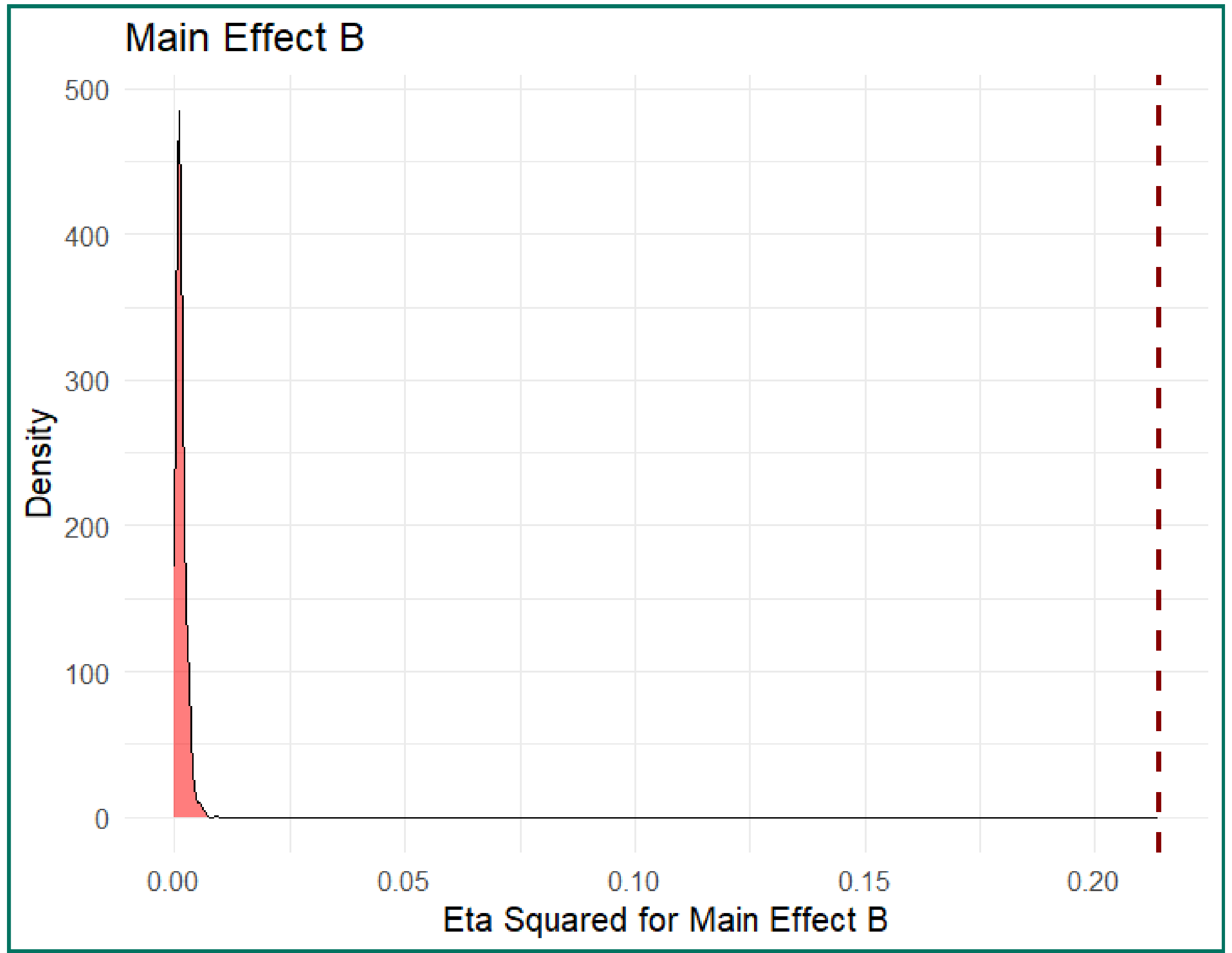
- **Homoscedasticity**
- **Absence of Outliers**

Permutation Results

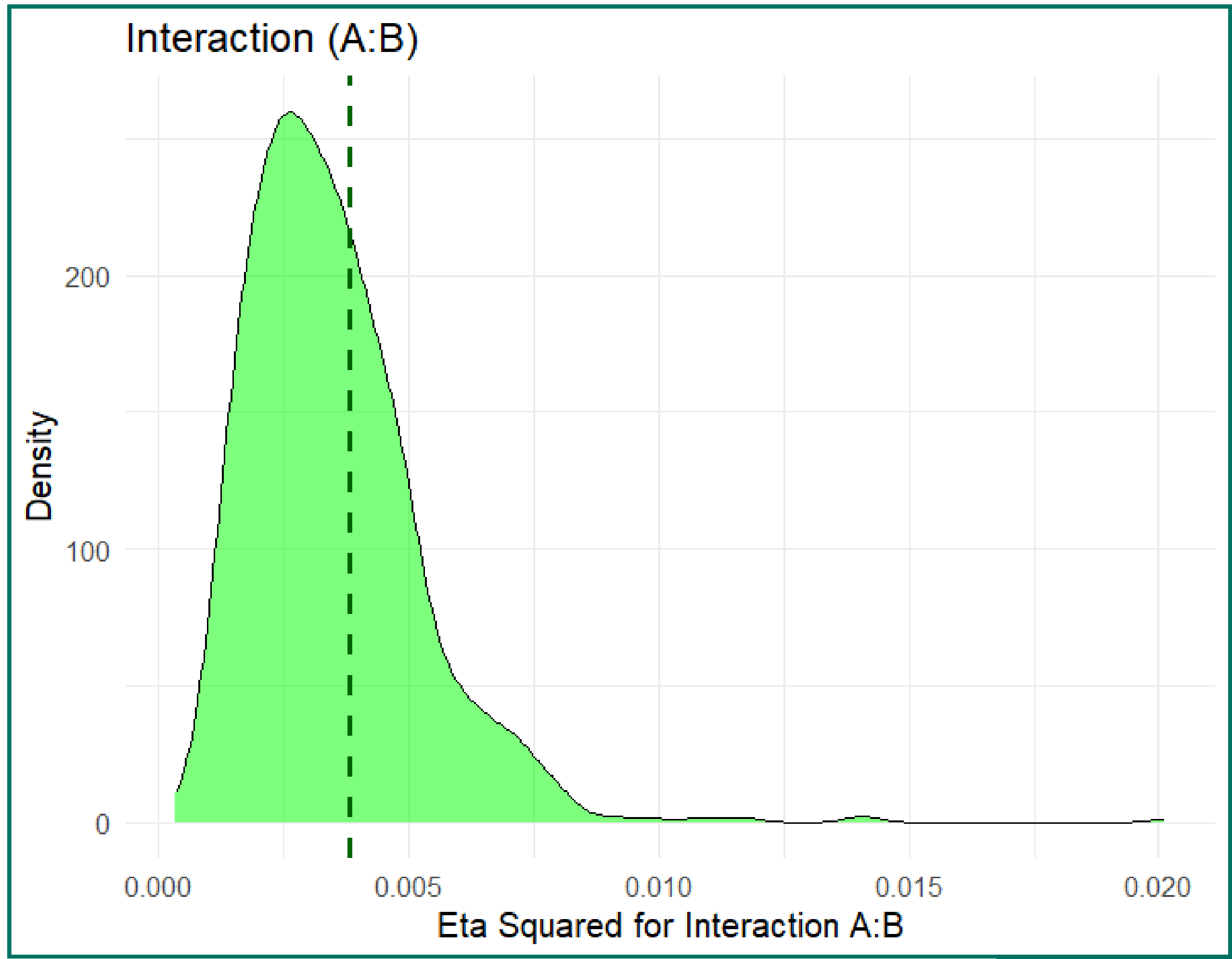
1 Main effect of A



2 Main effect of B



3 Interaction effect of A & B



Comparison with Parametric Two-Way ANOVA

OUR APPROACH

TERM	η^2	P value	Significance
A	0.0050	0.002	***
B	0.2138	0.000	***
A : B	0.0038	0.337	

TRADITIONAL ANOVA APPROACH

TERM	F value	P value	Significance
A	6.2557	0.0003	***
B	200.2953	<2.2e-16	***
A : B	1.4331	0.1591	

Interpretation

- Both methods yield similar results.
- Factor A and B have significant effects while interaction term is not significant at 5% significance level.

ART Limitations

ART APPROACH

TERM	F value	P value	Significance
A	4.2175	0.0055	**
B	283.9718	< 2.22 e-16	***
A : B	2.8747	0.0022	**

ART cannot be applied for original data since there are combinations with 0 occurrences, ART requires each combination to have at least 1 occurrence.
Hence, 'Po' level of factor B has to be removed in order to conduct ART analysis.

- Factor A, B and the interaction term are all significant at 5% significance level.

Data must be modified in order to compare all 3 methods !

Comparison for modified data

OUR APPROACH

TERM	η^2	P value	Significance
A	0.0052	0.001	***
B	0.2121	0.000	***
A : B	0.0037	0.276	

TRADITIONAL ANOVA APPROACH

TERM	F value	P value	Significance
A	6.4908	0.0002	**
B	264.2683	< 2.22 e-16	***
A : B	1.5469	0.1257	

ART IS MORE SENSITIVE TO SMALL EFFECTS AND INTERACTIONS, WHILE OUR METHOD AND TRADITIONAL ANOVA RELY ON MORE DIRECT ANALYSIS FOCUSING ON LARGER EFFECTS.

R PACKAGE : PERMOVA

- Link: https://github.com/TishaniSakalya/Permova

- Contents:
  - Functions
  - Documentation
  - Test cases
  - Vignette

TIPS FOR USING THE PACKAGE

- Ensure necessary libraries are installed.
- Assign the 2 factors as A and B.
- Assign dependent variable as Y.
- Data must be stored in a data frame.
- Refer to documentation for further use of functions and their results.

CONCLUSION

Summary

- Novel methodology
- Handle interactions
- Fewer assumptions
- Interpretable results
- Ideal for:
  - Small datasets
  - Balanced data
  - Violations of traditional ANOVA assumptions
  - Incomplete two-way factorial

Limitations

- Computational Feasibility
- Sample size dependence
- R package only applicable for 2 factors

Potential for Improvements

- Extend method for unbalanced design
- Extend R package for multiple factors
- Use of Parallel Computing or cloud-based resources

Group Members

Tishani Wijekoon
s16379

Lasandi Perera
s16388

Anupama Chandrapala
s16295

References

- **Making your first R package**
  - https://tinyheero.github.io/jekyll/update/2015/07/26/making-your-first-R-package.html
- **ARTool**
  - https://depts.washington.edu/acelab/proj/art/index.html
- **Nonparametric Tests of Interaction in Experimental Design by Shlomo S. Sawilowsky**
- **Effect sizes**
  - https://bookdown.org/a_shaker/STM1001_Topic_7/6-effect-sizes.html
- **Permutation test**
  - https://en.wikipedia.org/wiki/Permutation_test
- **Parallel Processing in R**
  - https://dept.stat.lsa.umich.edu/~jerrick/courses/stat701/notes/parallel.html
- **Stack Exchange Cross Validated-Discussions**
  - https://stats.stackexchange.com/
- **Nonparametric Two-Way ANOVA**
  - https://www.cfholbert.com/blog/nonparametric_two_way_anova/
- Scheirer, C. J., Ray, W. S., & Hare, N. (1976). **The analysis of ranked data derived from completely randomized factorial designs**. Biometrics, 32(2), 429-434.
  - https://doi.org/10.2307/2529511
- **Anderson, M. J., & Ter Braak, C. J. F. (n.d.). Randomization Tests with R**
- **Permutational Analysis of Variance - Wikipedia**

THANK YOU!

Any questions or comments?