A NOVEL APPROACH TO NON-PARAMETRIC TWO-WAY ANOVA : PERMOVA

AUTHOR

TISHANI WIJEKOON
Index: S16379
IS 3003 RESEARCH

01. THEORETICAL DEVELOPMENT

✅ Why? Traditional two-way ANOVA assumes normality, homoscedasticity—often violated in real-world data.
✅ Solution: Develop a bootstrap + permutation method for η^2 estimation and hypothesis testing.
✅ Key Idea: Uses resampling techniques instead of parametric assumptions.

ART ANOVA 🎯

- No normality needed
- Ranks & aligns data before running traditional ANOVA
- Interaction Effects: ✅ Yes
- Effect Size: Not directly estimated, ANOVA F-test used
- Computational Cost:⚡ Fast

Bootstrap-Permutation ANOVA 🔁

- Fully distribution-free
- Bootstraps effect sizes (η^2)
- Permutations for significance testing
- Interaction Effects: ✅ Yes
- Effect Size: Provides bootstrap CI for η^2
- Computational Cost: ⏳ Slower, requires many resamples

02. METHODOLOGY

✅ Step 1: Bootstrap Sampling
- Resample data with replacement to estimate the distribution of η^2.
- Provides confidence intervals without parametric assumptions.

✅ Step 2: Permutation Testing
- Shuffle factor labels to generate a null distribution of η^2.
- Compute p-values based on permutation-based significance testing.

TEST STATISTIC

Measures proportion of variance explained

Effect Size

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

HYPOTHESIS

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

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

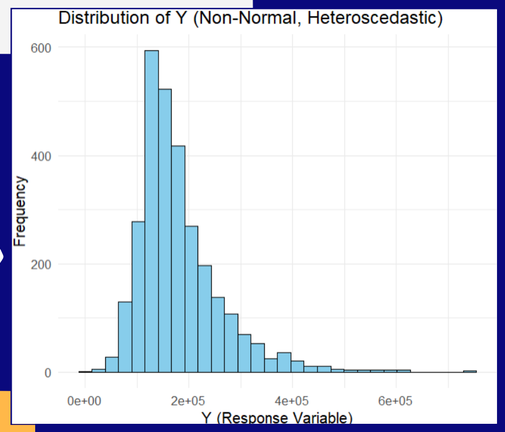
04. REAL DATA APPLICATION

A dataset of 2930 observations, obtained from Kaggle was used. Factor A Season has 4 levels, Factor B Heating Quality has 5 levels. Link here

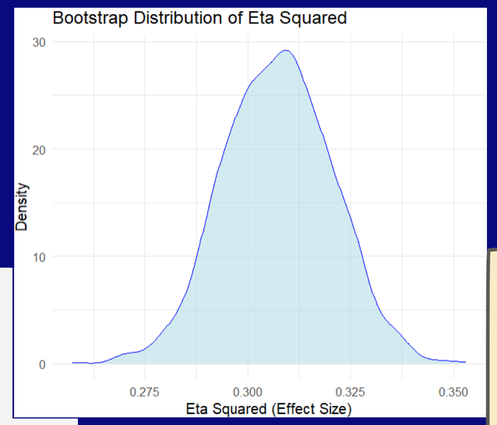
DATA CATEGORY TABLE

	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

DISTRIBUTION OF ORIGINAL DATA



BOOTSTRAPPING



Bootstrap Type : Winsorized
mean_Eta: 0.3073
ci_lower: 0.2812
ci_upper: 0.3333
se_eta: 0.0131

RESULTS OF REAL DATA

OUR APPROACH

Term	η^2	P value	Sig.
A	0.0050	0.002	***
B	0.2138	0.000	***
A*B	0.0038	0.337	

ANOVA APPROACH

Term	F value	P value	Sig.
A	0.3004	0.000	***
B	0.3407	0.000	***
A*B	0.0459	0.003	

03. SIMULATION RESULTS

This synthetic dataset is of 200 sample size with interaction effects, non-normal residuals, and heteroscedasticity is ideal for testing non-parametric methods.

OUR APPROACH

Term	η^2	P value	Sig.
A	0.3004	0.000	***
B	0.3407	0.000	***
A*B	0.0459	0.003	**

ART APPROACH

Term	F value	P value	Sig.
A	325.450	2.22e-16	***
B	323.403	2.22e-16	***
A*B	49.532	3.20e-11	***

ART approach cannot be used for original data since, there are is 0 occurrences of combinations. ART requires each combination to have at least 1 observation. Therefore, the specific level of Factor B was removed to conduct ART analysis.

ART APPROACH

Term	F value	P value	Sig.
A	4.2175	0.0055	**
B	283.9718	<2e-16	***
A*B	2.8747	0.0022	**