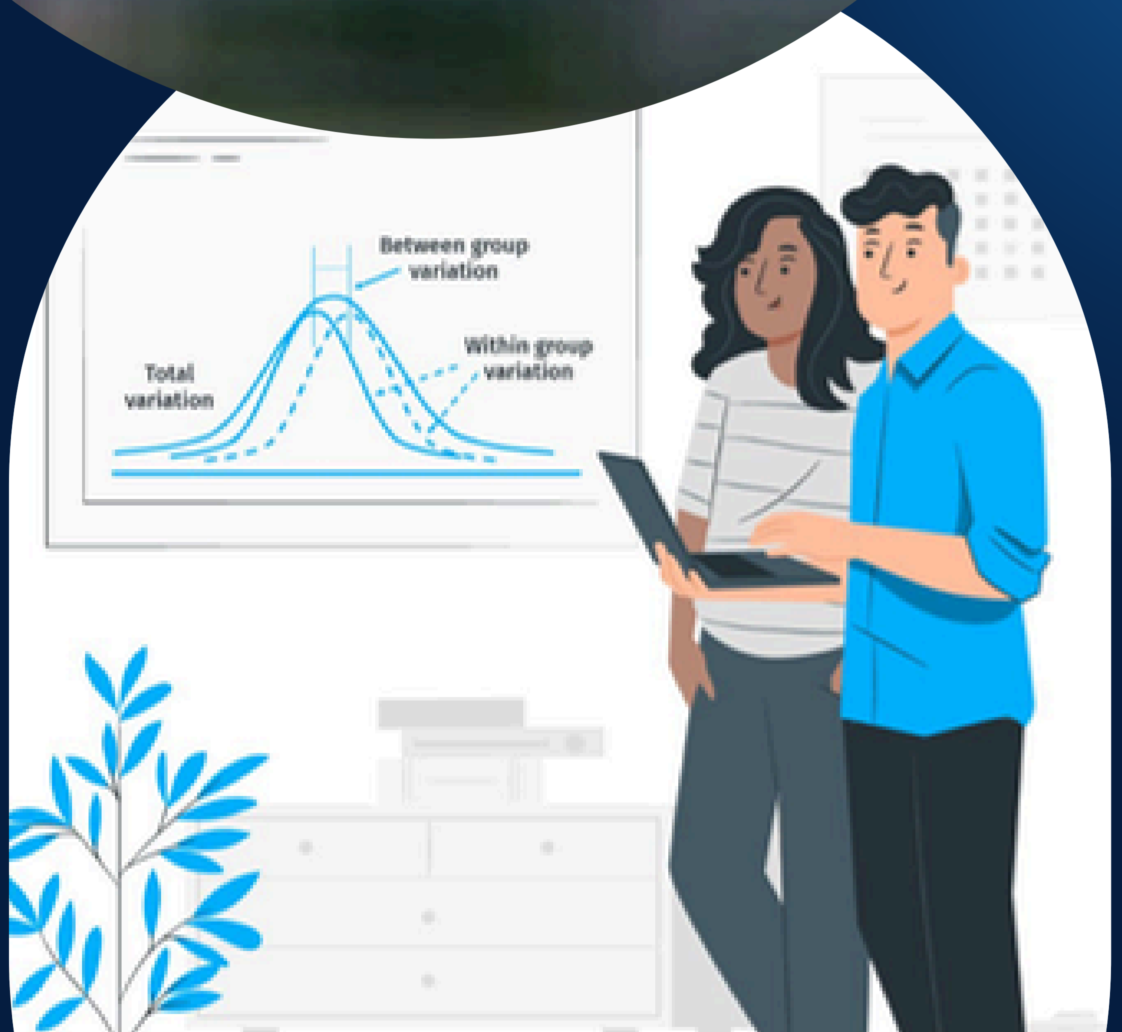
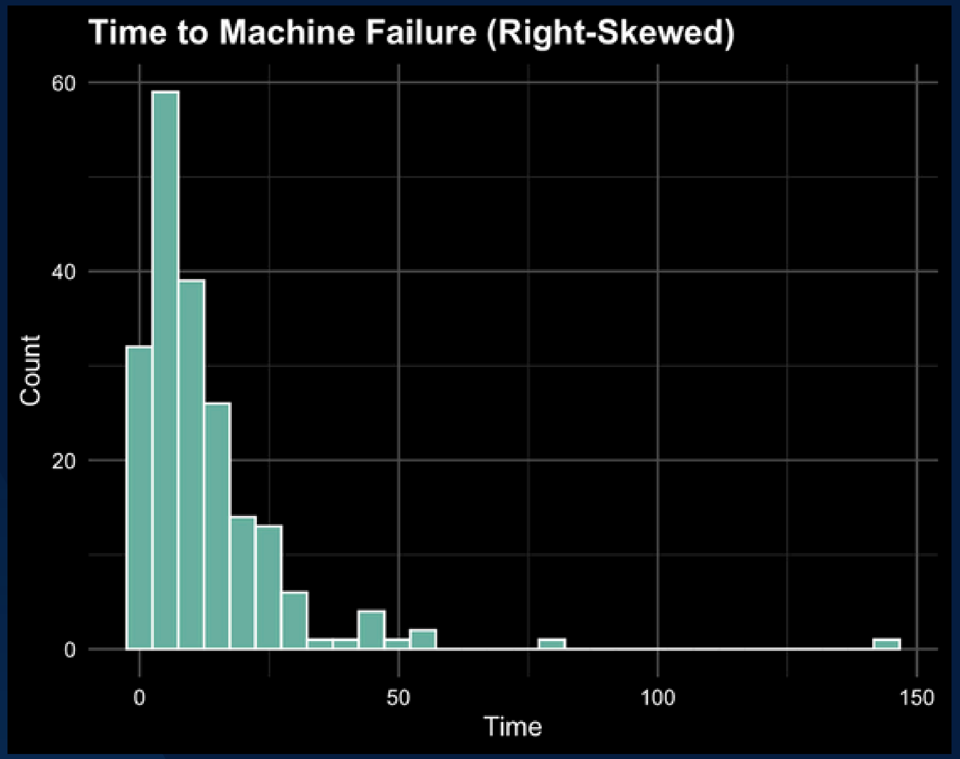
NON PARAMETRIC TWO WAY ANOVA

Presented by: Group 2

- Poornima Tharangani - s16368
- Pasindu madusanka-s16333
- Vayani Kavindya - s16322



WHY NON-PARAMETRIC TWO WAY ANOVA

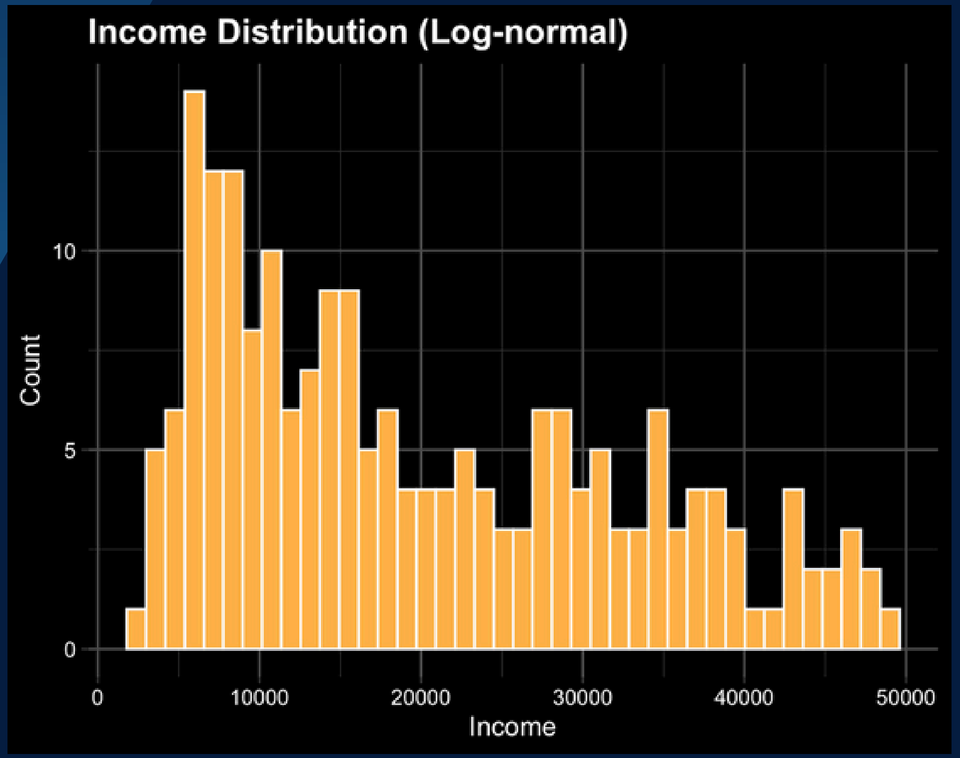


Time Until a Machine Fails

measure time-to-failure.
two machine models :- A, B
two stress levels :- Low, High

Customer Ratings

Customer ratings across
product category :- Electronics, Clothing
review type :- Verified, Unverified



Income level

income across
education level:- High School, College
job sector:- Tech, Service

Why not Parametric two way ANOVA

PARAMETRIC ANOVA ASSUMPTIONS

Y_ijk = μ + αi + ßi + (αß)ij + £_ijk

Y_ijk ~ N(μij , ∂^2)

Dependent variable is Normally distributed with

- mean μij
- Constant variance ∂^2

Y_ijk s are independent

£_ijk ~ N(0, ∂^2)

Errors are Normally distributed with

- mean 0
- Constant variance ∂^2

Errors are independent and identically distributed

Our focus is Normality Assumption

Existing non parametric approaches

Rank based methods:

Aligned Rank Transformation

- **Handles main and interaction effects without parametric assumptions.**
- **Aligns and ranks data for each effect before ANOVA.**
- **Less effective with tied or highly skewed data.**

Regression based methods

- **Use flexible function estimation to capture nonlinear effects.**
- **Suitable for high-dimensional or complex structured data.**
- **Do not rely on normality or homoscedasticity assumptions.**
- **Involve complex mathematical concepts and implementation.**

Smoothing Spline ANOVA

- Models main and interaction effects using smoothing splines.

Kernel ANOVA

- Uses kernel functions to estimate additive and interaction effects.

MATHEMATICAL FRAMEWORK

Functional ANOVA

F-ANOVA extends the traditional ANOVA method to analyze functional data by decomposing the functional response into components.

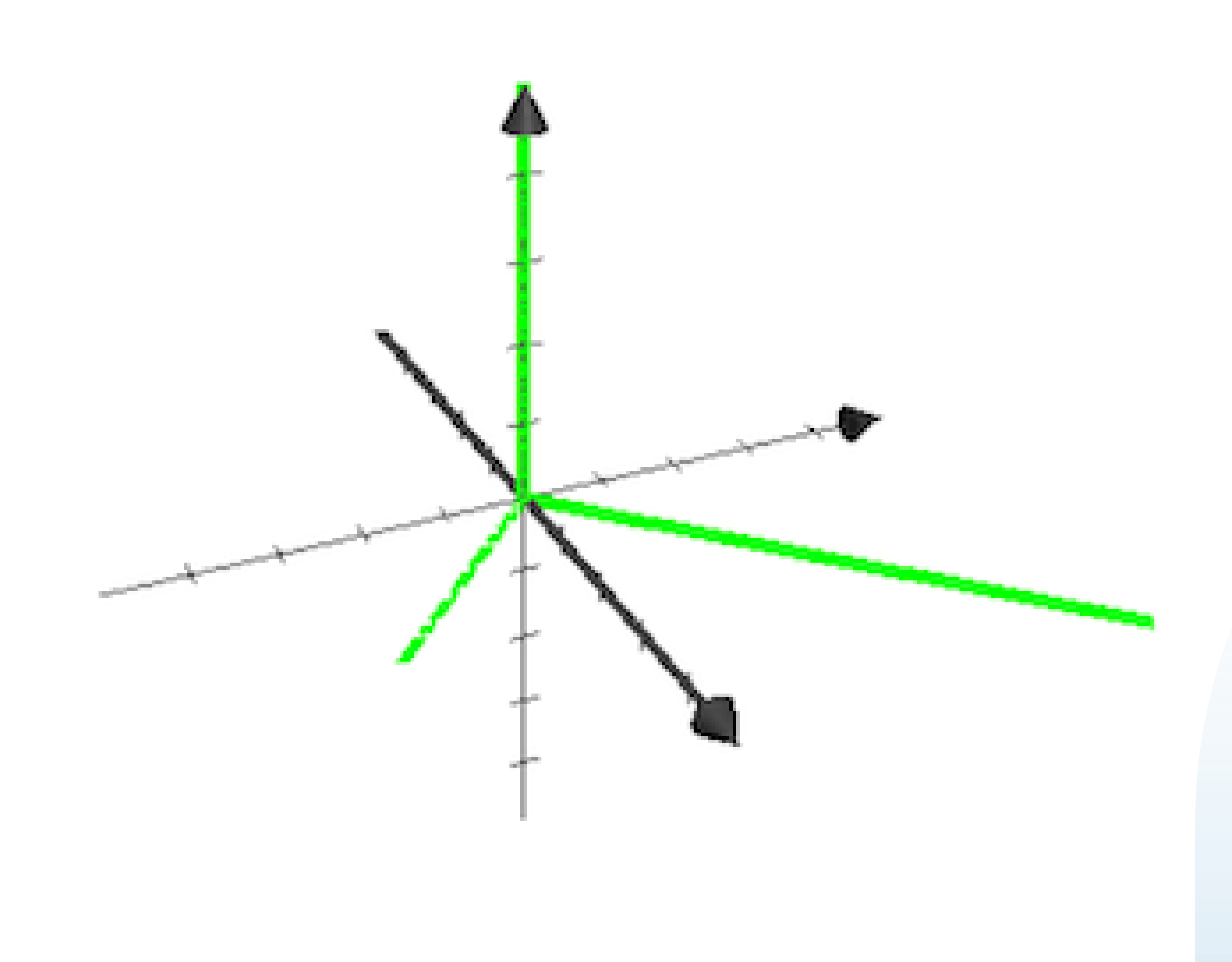
$$F(X) = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{i \neq j} f_{ij}(X_i, X_j) + \dots,$$

In Our Method we used fANOVA to decompose the Response function into ;

- Main Effects
- Interactions

WILL EFFECTS BE PURE ???

Othogonality by Purification



Why Othogonality is need ?

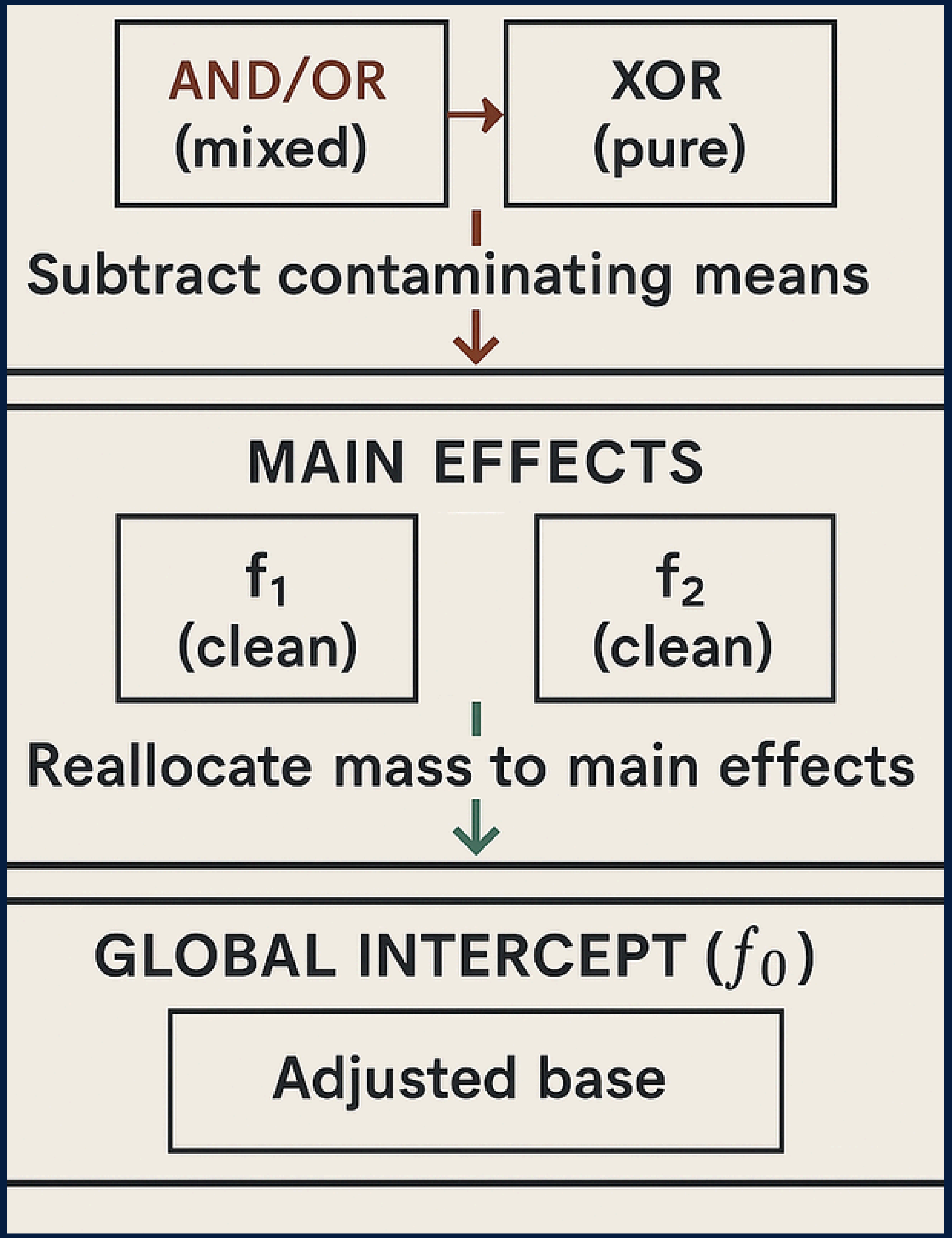
- Clear interpretation: How much variation each factor is truly contributing.
- Unbiased estimates: One factor's estimate isn't contaminated by the others.

The Models we estimate have an identifiability challenge

Effects can be freely moved between main effects and interaction effects

to overcome that we used **MASS MOVING PURIFICATION**

Mass moving algorithm



- It is an algorithm proposed by Lengerich (2020) to decompose a model's predictions into pure, non-overlapping components (main effects and interactions) by iteratively removing contaminating effects from higher-order interaction terms.
- It defined interaction effects as variance which can not be explained by main effects
- **Mass-Moving**: Redistribute "mass" (variance contributions) from higher-order interactions to lower-order terms until all effects are orthogonal.

Testing for Statistical Significance

Freedman-Lane Permutation

- generates a null distribution for the test statistic by permuting the residuals of a reduced model (excluding the effect of interest)
- Refit the full model to the permuted outcomes.
  (New Response)

Testing for Statistical Significance

$$H_0, H_1$$

Hypothesis

Testing Main Effects：

H0：The main effects do not have a significant impact on the response. (gj(xj)=0)

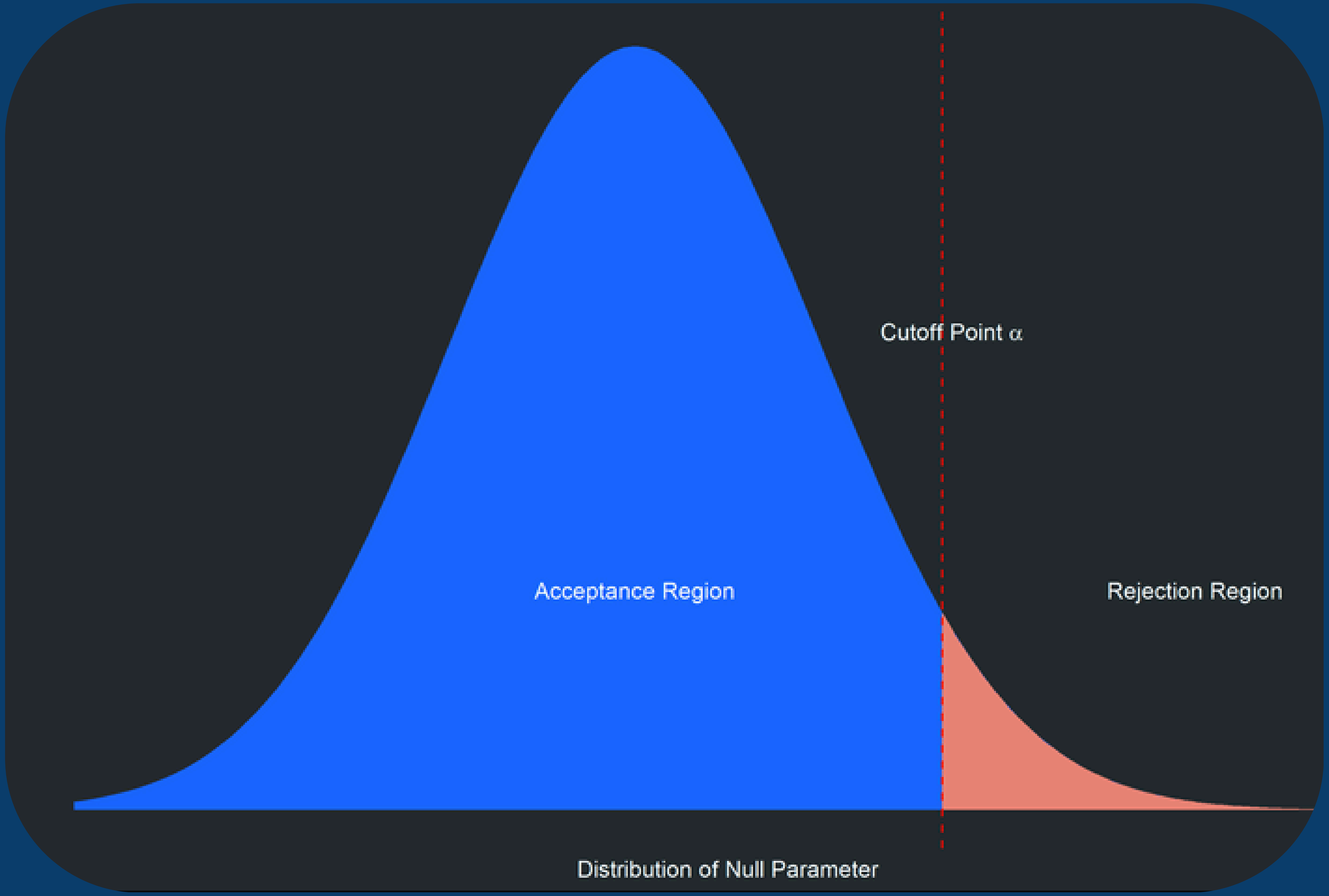
Ha：The main effects have a significant impact on the response

Testing interaction effect：

H0: The interaction effect do not have a significant impact on the response. (gjk(xj,xk)=0)

Ha：The interaction effects have a significant impact on the response

Testing for Statistical Significance



Null Distribution of R square

Sampling Distribution

This method constructs an empirical null distribution of the test statistic by assuming the null hypothesis is true.

Test Statistic

R square

METHODOLOGY

METHODOLOGY

- Categorical variables are encoded, and interaction terms are created as pairwise products of main effects

Variable 1	Variable 2
level 1	level 1
level 2	level 1
level 1	level 2
level 2	level 2

Variable 1	Variable 2	Interaction
0	0	0
1	0	0
0	1	0
1	1	1

- Need to keep 20% of the data as validation set

- Get the global mean

$$g_0(x) = \frac{1}{n} \sum y_i$$

Estimating main effects

Select one Main effect let's say "s"

Tm is the mth effect estimating tree added to the model in a one iteration

Σ [Tm(xs)]

- Estimated effect when x=0
- Estimated effect when x=1

How to fit Tm

- Trees(Tm) are fitted using the residuals(zi,m)
- In this case only 2 categorical variables exist.Therefore,

$$z_{i,m} = y_i - g_{m-1}(x_i)$$

Zi,m = Residual

Variable s	Residual(Zi,m)
0	Z1,m
1	Z2,m
1	Z3,m

Estimating main effects

How to fit Tm

Variable s	Residual(Zi,m)
0	Z1,m
1	Z2,m
1	Z3,m

Tm(xs)
- Mean of Zi,m where variable 1 is 0
- Mean of Zi,m where variable 1 is 1

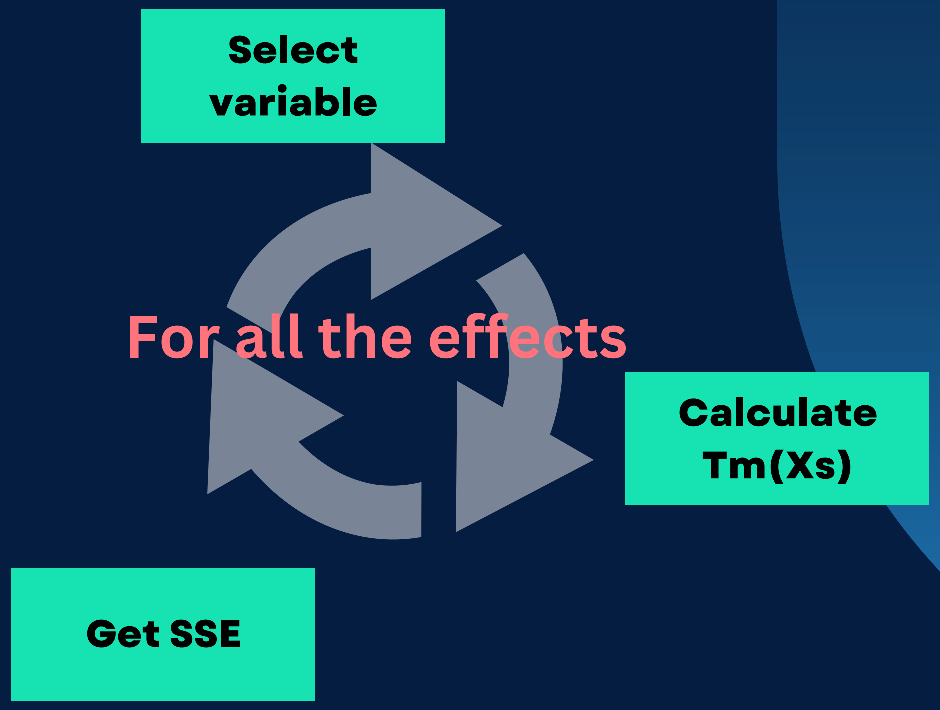
- for that tree Calculate SSEs

$$SSE_s = \sum_{i=1}^{n} \left(z_{i,m} - T_m(\boldsymbol{x}_{i,s}) \right)^2$$

Now select other effects and repreat the above process
Then we have SSE values for all the main effects
Effect with min SSEs(s*) add that to the model

$$g_m(\boldsymbol{x}) = g_{m-1}(\boldsymbol{x}) + \lambda T_m(\boldsymbol{x}_{s*})$$

(The learning rate λ controls updates to prevent overfitting)



Estimating main effects

Now we have a new model $g_m(\boldsymbol{x}) = g_{m-1}(\boldsymbol{x}) + \lambda T_m(\boldsymbol{x}_{S*})$

Using that model, calculate the validation set loss(SSE)

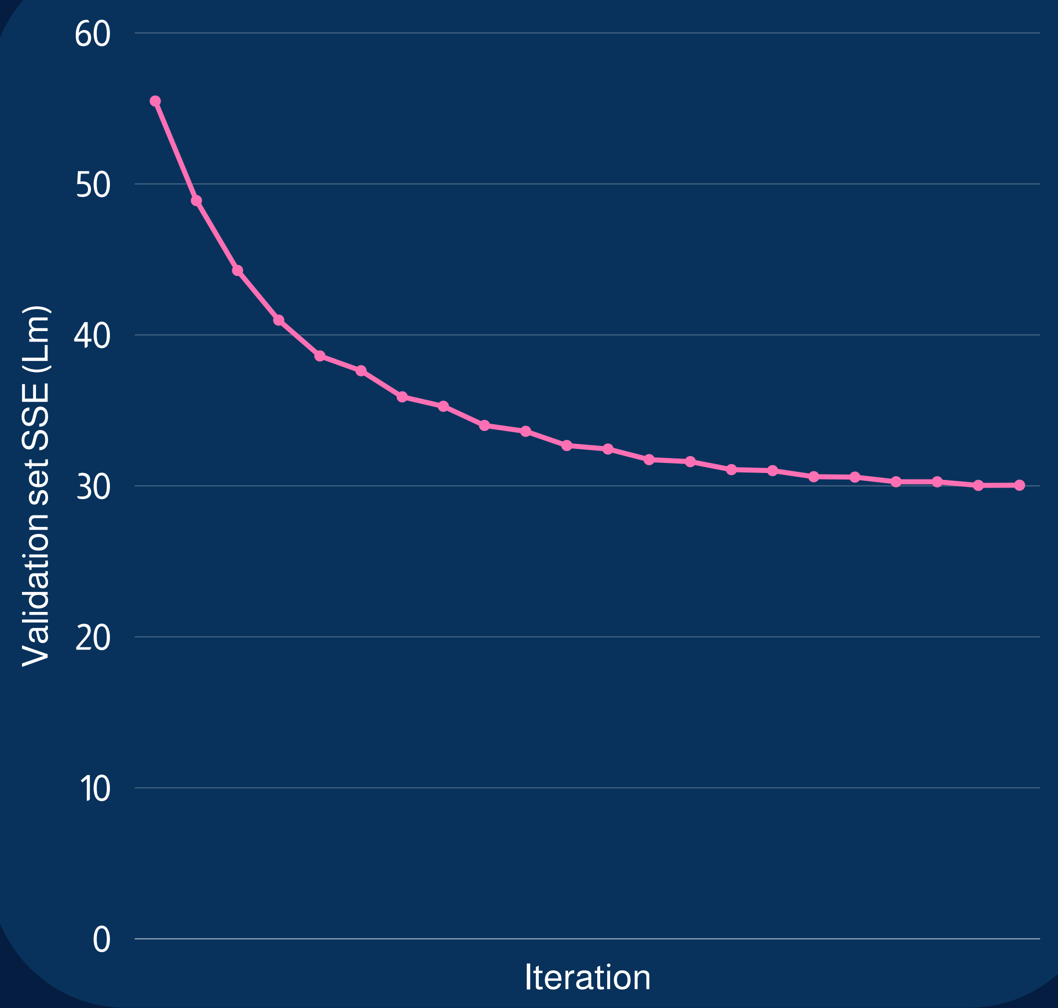
$$L_m = \frac{1}{n'} \sum_{i=1}^{n'} \ell(y_i, g_m(\boldsymbol{x}_i))$$

With the new model Residual(Zi,m) can be calculated again

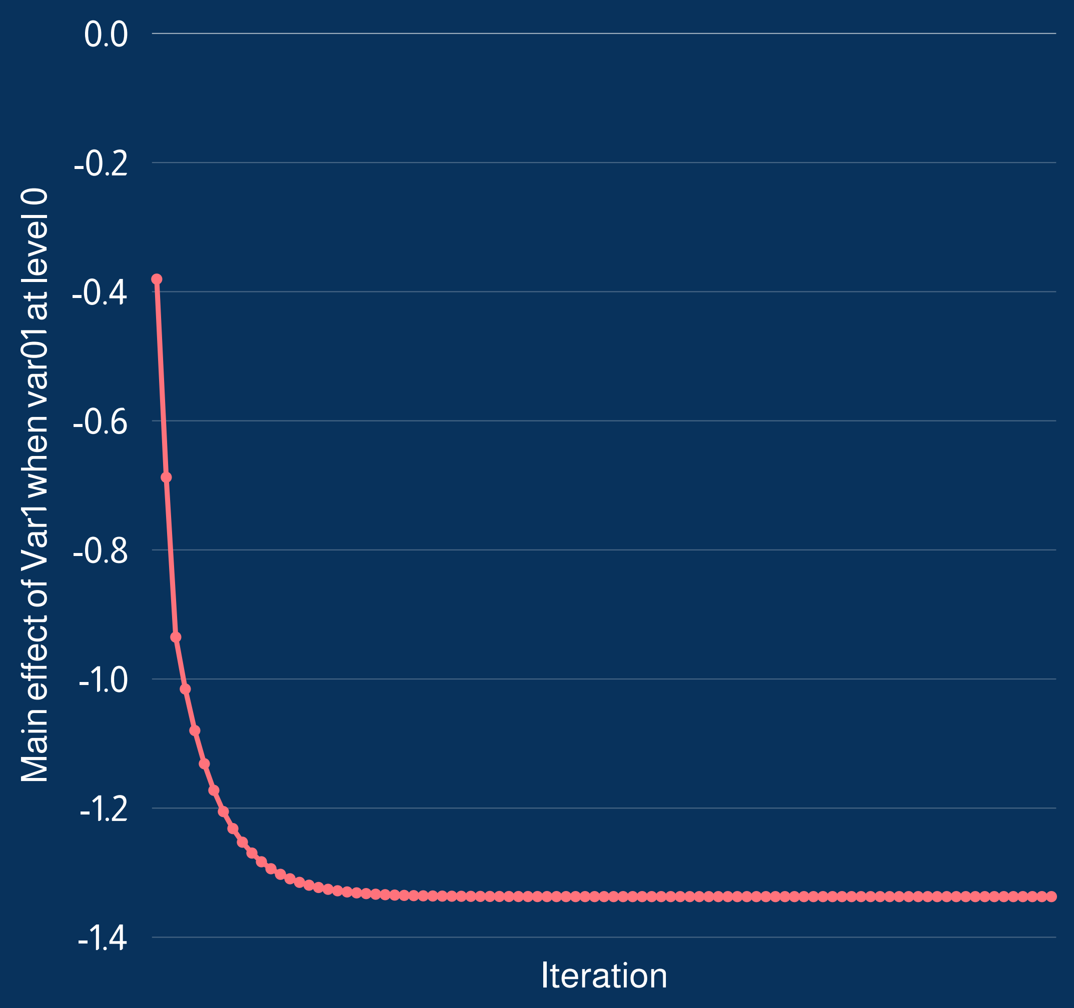
$$z_{i,m} = y_i - g_{m-1}(x_i)$$

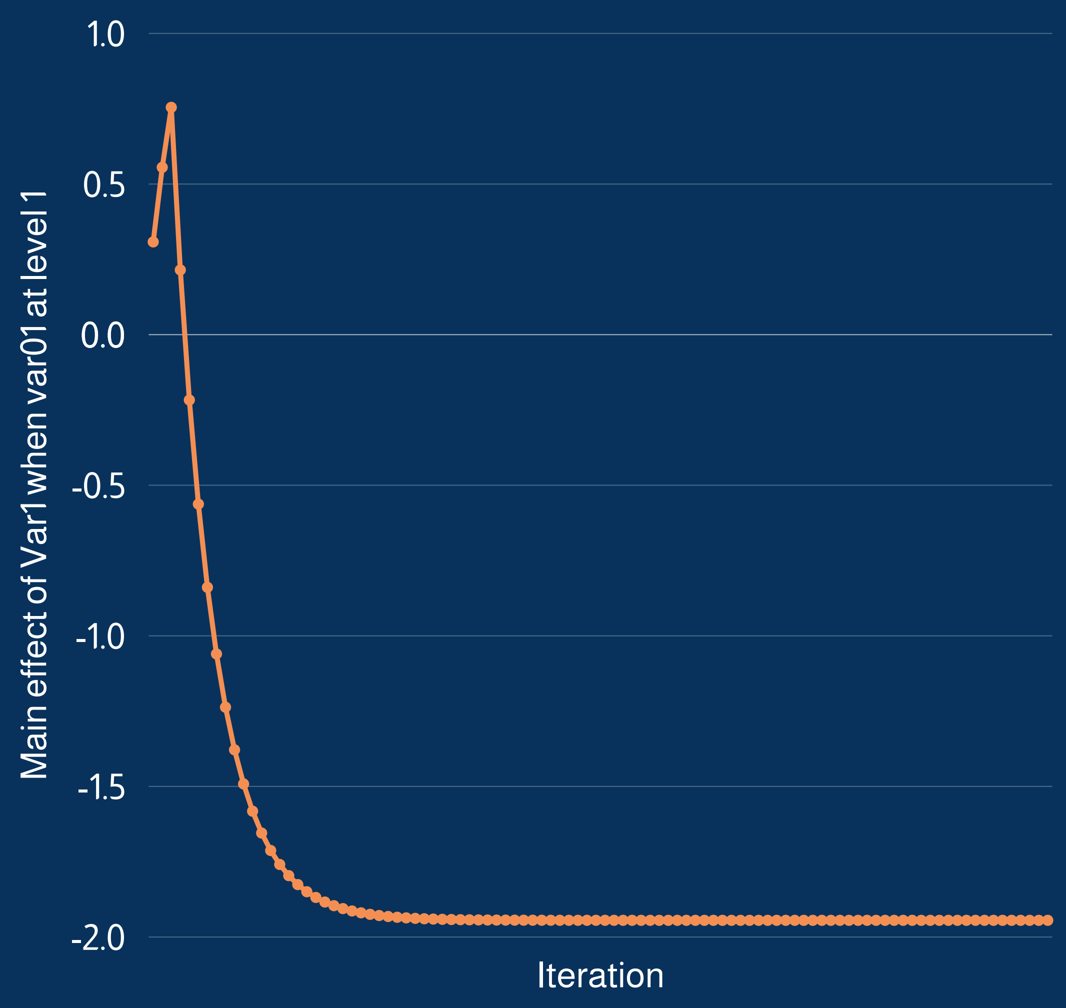
- Now the whole above process should be repeated again using this residual
- selecting effects up to getting the Lm

- The Lm values will be monitoted
- The above process repeat untill no improvement detected in Lm



How Main effects converge during the above process





Graphs for two levels of variable one

Estimating Interaction effects

The previous procedure is for the main effects

For interaction effects, the same procedure can be followed, but the mass moving alagorithm will be used to keep orthogonality

MASS MOVING PURIFICATION

- previously, we have assumed interaction effect as an AND operation but it also can be OR.
- After applying mass moving algorithm effects will be orthogonal to the main effects and XOR

Variable 1	Variable 2	Interaction	residual
0	0	0	Z1,m
1	0	0	Z2,m
0	1	0	Z3,m
1	1	1	Z4,m

calculate
mean of residuals where Var1=0 and Var2=0
mean of residuals where Var1=0 and Var2=1
mean of residuals where Var1=1 and Var2=0
mean of residuals where Var1=1 and Var2=1

Estimating Interaction effects

Method

Var1/Var2	0	1
0	a	b
1	c	d

a- [mean of residuals where Var1=0 and Var2=0]*learning rate
b- [mean of residuals where Var1=0 and Var2=1]*learning rate
c- [mean of residuals where Var1=1 and Var2=0]*learning rate
d- [mean of residuals where Var1=1 and Var2=1]*learning rate

Mean of values in first row

Mean of values in second row

Each row value - Mean of the row & Main effect of Var 1 + Mean of the rows

Var1/Var2	0	1
0	a	b
1	c	d

Var1/Var2	0	1
0	(a-b)/2	-(a-b)/2
1	(c-d)/2	-(c-d)/2

Mean of values in first column

Mean of values in second column

Each column value - Mean of the column & Main effect of Var 2 + Mean of the columns

Var1/Var2	0	1
0	(a-b)/2	-(a-b)/2
1	(c-d)/2	-(c-d)/2

Var1/Var2	0	1
0	[(a-b)-(c-d)]/4	-[(a-b)-(c-d)]/4
1	-[(a-b)-(c-d)]/4	[(a-b)-(c-d)]/4

Final Interaction Table

Estimating Interaction effects

Variable 1	Variable 2	Interaction
0	0	0
1	0	1
0	1	1
1	1	0

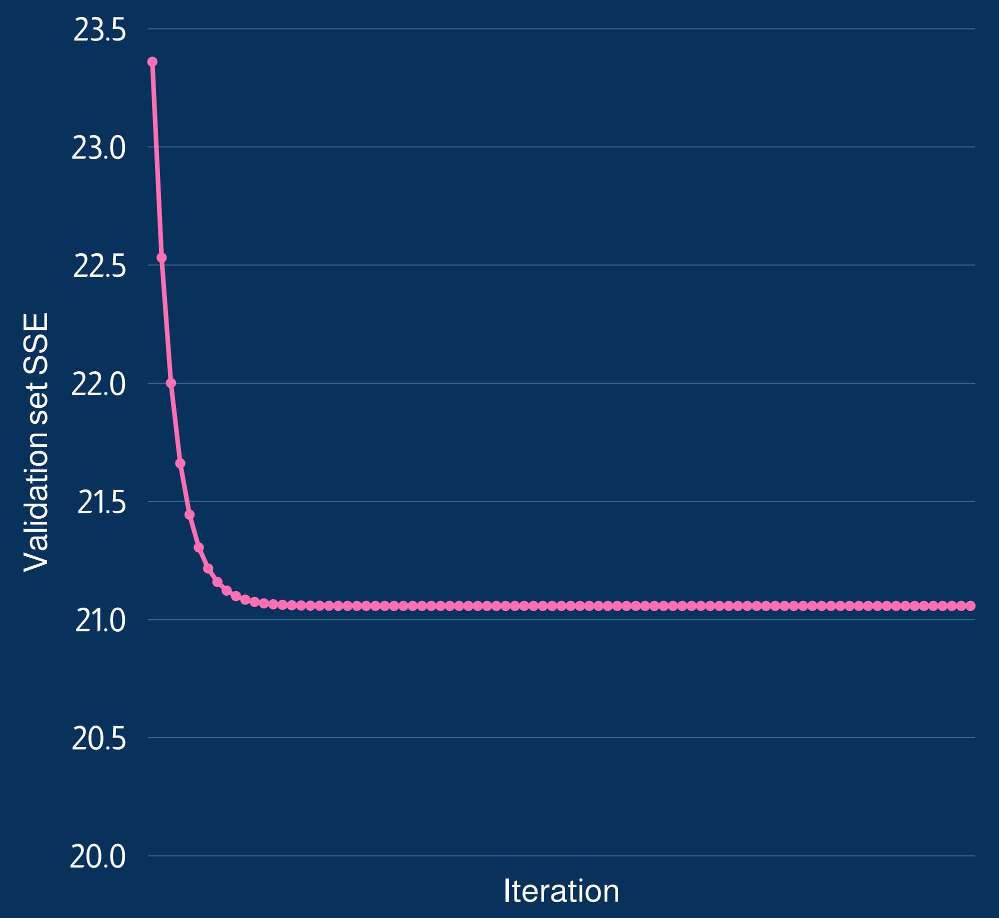
Var1/Var2	0	1
0	[(a-b)-(c-d)]/4	-[(a-b)-(c-d)]/4
1	-[(a-b)-(c-d)]/4	[(a-b)-(c-d)]/4

- After applying the mass moving algorithm effects will be orthogonal and XOR

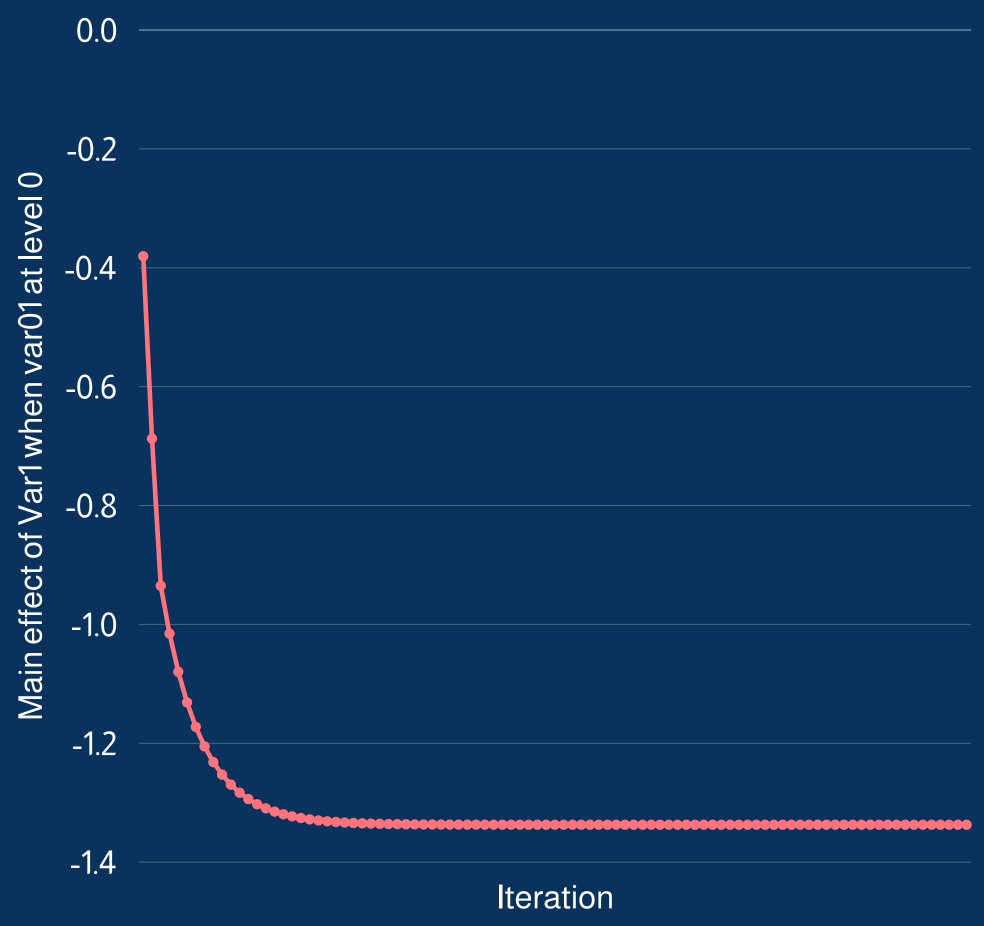
- then the Interaction term will be included to the model

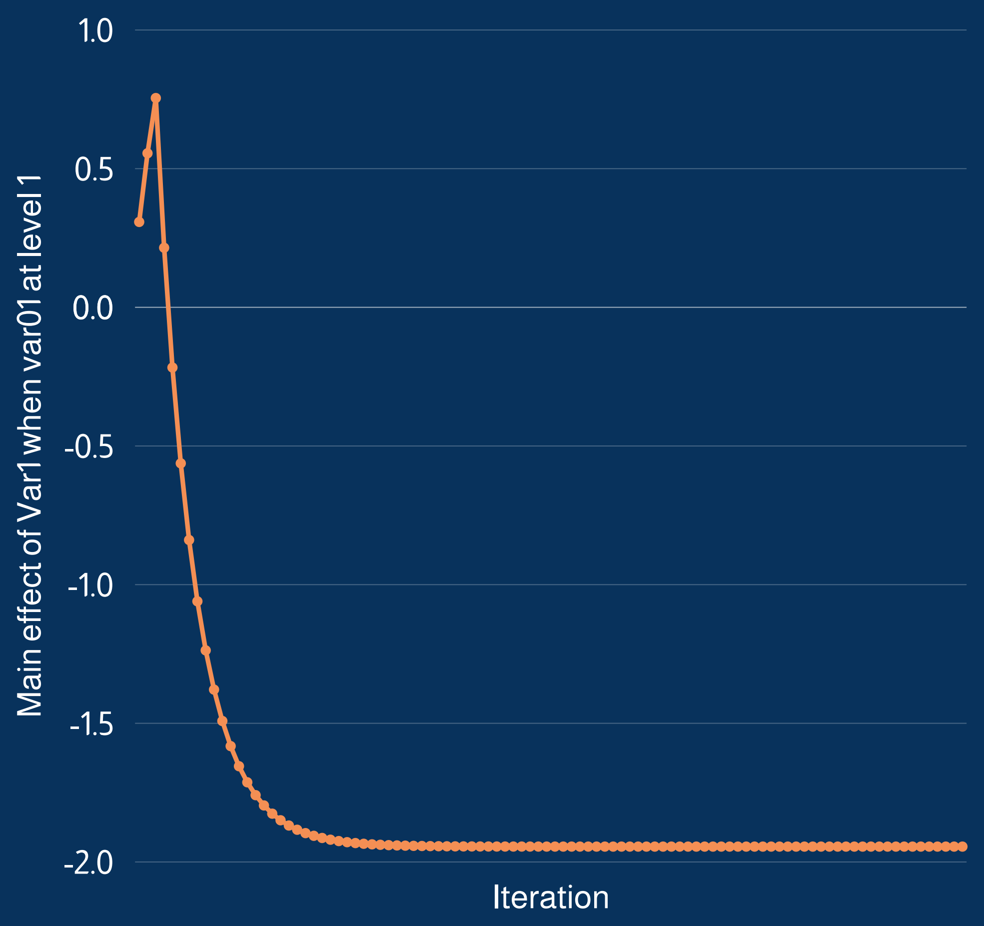
$$\text{Tm(x)\_interaction} \begin{cases} [(a-b)-(c-d)]/4 \text{ when var1=var2} \\ -[(a-b)-(c-d)]/4 \text{ when var1≠var2} \end{cases}$$

- In the previous step, interaction effects, and main effects both are changed
- above process also will be repeated until no further improvement is identified



Interaction effects also converge same as the main effects during the above process





Example Final model

Level	Factor A	Factor B	FactorA_FactorB	Global_mean
0	-2.247049	-7.257762	0.001915252	52.11115
1	1.685287	6.854553	-0.001915252	-

Permutation Test for Statistical Significance

01 Fit the full model with original responses and calculate Observed R square

Y = F(X) + £

F(X) = f0 + f1(X) + f2(X) + f12(X)

Calculate R^2 value: R^2

02 By removing a selected variable fit a reduced model and get the residuals

F(X) = f0 + f2(X) + f12(X) <--- f1(X) removed

using F(X) predict Y: Y_pred

£ = Y - Y_pred

03 Shuffle residuals of reduced model and add to Y_pred

Shuffled residual: £

New response: Y = Y_pred + £

Permutation Test for Statistical Significance

04 Refit the full model using new responses and the permuted R square values are obtained

$Y = F(X) + £$

$F(X) = f0 + f1(X) + f2(X) + f12(X)$ <--- new full model

Calculate R^2 value: R^2

If the effect is significant, shuffling the residual should lead to a significant drop in R^2, confirming its contribution.

05 Repeat the above step 03 and step 04 N times, and Get the P value

P = [number of times R^2 > R^2] / N

If P<0.05, the effect of Var1 is significant at a 5% significant level

SIMULATED RESULTS

Simulated Results

- Categorical variables:
  - A (A1,A2)
  - B (B1,B2)
- Response variable:
  - Influenced by exponential noise, creating a non-normal distribution.

Level	Factor A	Factor B	FactorA_FactorB	Global_mean
0	-2.247049	-7.257762	0.001915252	52.11115
1	1.685287	6.854553	-0.001915252	-

Respective P-values

Variables	p-values
A	0.06930693
B	0.00990099
Interaction	0.5049505

The data was simulated to have a significant effect by Factor B

Simulated Results

Categorical variables:

- A (A1,A2)
- B (B1,B2)

Level	Factor A	Factor B	FactorA_FactorB	Global_mean
0	-0.4775220	-7.047752	6.661338e-16	47.21881
1	0.3581415	6.656211	-6.661338e-16	-

Response variable:

- follows a different pattern:
- If A = A1, the standard deviation is 1.
- If A = A2, the standard deviation is 5.

Respective P-values

Variables	p-values
A	0.4851485
B	0.00990099
Interaction	0.1089109

The data was simulated to have a significant effect by Factor B

REAL DATA APPLICATIONS

Real Data Applications

Medical Cost Personal Dataset

Categorical variables:

- sex (Male-0, Female-1)
- smoker (yes-1, no-0)

The response variable:

- medical charges

Response is highly skewed and non-normal (Shapiro-Wilk $p < 2.2e-16$)

Level	Sex	Smoker	Sexfe_Smoker	Global_mean
0	46.95336	-4711.766	-24.7637	13249.6
1	-48.37927	18258.093	24.7637	-

Respective p-values

	P-values
Sex	0.8431373
Smoker	0.01960784
Interaction	0.8235294

Students Performance in Exams dataset

Categorical variables:

- Gender (Male-0 and Female-1)
- Test.preparation.course (none-1, completed-0)

Level	gender	test preparation	gender_test_prep	Global_mean
0	-2.483320	2.662016	0.009910888	66.68286
1	2.752595	-1.681600	-0.009910888	-

The response variable:

- Math score

Respective p-values

	P-values
gender	0.01960784
test preparetion	0.05882353
Interaction	0.7058824

Response is bit skewed and vary from non-normality
(Shapiro-Wilk p= 0.0001455)

Power Analysis

This power analysis examined detectability of main effects and interactions across sample sizes (n=60, 100, 150)

Two categorical factors:

- Factor A (moderate effect)
- Factor B (strong effect)

sample size	power_factorA	power_factorB	Power_Interaction
60	0	1	0.75
100	0.6363636	1	0.8636364
150	0.7837838	1	0.8648649

50 simulations (due to computational cost)

Results:

- Factor A

Required larger samples (64% power at n=100; 78% at n=150) failed detection at n=60 (0% power).

- Factor B

high observed power (100%) -likely overestimate true power due to low iterations.

- Interaction

High observed power (above 75%)

Comparison

	Parametric Two way ANOVA	This Approch
Normality	Assumes that the residuals are normally distributed.	Does not assume normality.
Homogeneity of Variance	Assumes the homogeneity of variance of the residuals.	Does not assume equal variances.
Observation are independent	Observations are assumed to be independent.	Observations are assumed to be independent.
Sampling distribution	Significance is assessed by using the F distribution.	Significance is assessed by comparing the observed statistic to its distribution under random permutations.
Computational efficiency	Computationally efficient and fast, especially for moderate samples and few factors.	The boosting process and multiple permutations increase computational demands.
Sum to 0 constraints	Ensures identifiability by enforcing sum-to-zero constraints for factor effects.	Sum-to-zero constraints may not be naturally imposed.

Implementation Challenges...

Computational efficiency & Performance Tradeoff

More permutations → more accurate p-value estimates

But also → more models, more Time

Good performance— just implement a few more hours... or days

Too complex for Pen and Papers

Our method involves residual reshuffling, multiple rounds of model training, and careful R^2 comparisons

Highly correlated predictors

Removing one may not affect to R square much so permutation tests may underestimate its true importance

Conclusion

- We proposed a Regression-based, interpretable framework for nonparametric ANOVA.
- Using a permutation test, we assess effect significance without strong assumptions.
- Simulation Results confirm its accuracy
- Real-world data supports its reliability— making it a robust and practical tool for modern analysis.

References

- GAMI-Net: An Explainable Neural Network based on Generalized Additive Models with Structured Interactions. Zebin Yang Aijun Zhang and Agus Sudjianto arXiv:2003.07132v2 [stat.ML] 2 Jun 2021
- . Interpretable Machine Learning based on Functional ANOVA Framework: Algorithms and Comparisons. Linwei Hu, Vijayan N. Nair, Agus Sudjianto, Aijun Zhang, and Jie Chen Corporate Model Risk, Wells Fargo, USA May 23, 2023.
- Purifying Interaction Effects with the Functional ANOVA: AnEfficient Algorithm for Recovering Identifiable Additive Models. Benjamin Lengerich, Sarah Tan, Chun-Hao Chang, Giles Hooker, and Rich Caruana
- . Permutation tests for linear models. Marti J.Anderson and John Robinson University of Sydney.Aust.N.Z.J.Stat. 43(1), 2001, 75-88

https://brainder.org/2020/05/19/simplifying-freedman-lane/

THANK YOU

GROUP 02

Presented by:

- Poornima Tharangani - s16368
- Pasindu madusanka-s16333
- Vayani Kavindya - s16322