

Supplementary Material for Principal Interactions Analysis for Repeated Measures Data: Application to Gene-Gene, Gene-Environment Interactions

BHRAMAR MUKHERJEE^{1,*}, YI-AN KO¹, TYLER VANDERWEELE², ANINDYA ROY³,
SUNG KYUN PARK⁴, AND JINBO CHEN⁵

¹ Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109

² Department of Epidemiology and Biostatistics, Harvard University, Boston, MA 02115

³ Department of Statistics, UMBC, Baltimore, MD 21250

⁴ Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109

⁵ Department of Biostatistics, University of Pennsylvania, Philadelphia, PA 19104

**email:* bhramar@umich.edu

A.1: Tables of analysis of variance corresponding to four interaction models

Table 1: ANOVA Table corresponding to Tukey's one d.f. nonadditivity test

Source	SS	df	F
Mean	$SSm = IJ\hat{\mu}^2$	1	
Row Main	$SS_R = J \sum_i \hat{R}_i^2$	$I - 1$	MS_R/MSE
Column Main	$SS_C = I \sum_j \hat{C}_j^2$	$J - 1$	MS_C/MSE
Interaction	$SS_u = \hat{\theta}^2 \sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2$	1	MS_u/MSE
Error	$SSE = SST - SSm - SS_R - SS_C - SS_u$	$(I - 1)(J - 1) - 1$	
Total	$SST = \sum_i \sum_j y_{ij}^2$	IJ	

Table 2: ANOVA Table corresponding to Mandel's column-regression model

Source	SS	df	F
Mean	$SSm = IJ\hat{\mu}^2$	1	
Row Main	$SS_R = J \sum_i \hat{R}_i^2$	$I - 1$	MS_R/MS_{res}
Column Main	$SS_C = I \sum_j \hat{C}_j^2$	$J - 1$	MS_C/MS_{res}
Column Slopes	$SS_s = \sum_i \sum_j \hat{\lambda}_i^2 \hat{C}_j^2$	$I - 1$	MS_s/MS_{res}
Residuals	$SS_{res} = SST - SSm - SS_R - SS_C - SS_s$	$(I - 1)(J - 1) - (I - 1)$	
Total	$SST = \sum_i \sum_j y_{ij}^2$	IJ	

Table 3: ANOVA Table corresponding to Gollob's F test for the AMMI model with $M < (I - 1)$ components, $I < J$.

Source	SS	df	F
Mean	$SSm = IJ\hat{\mu}^2$	1	
Row Main	$SS_R = J\sum_i \hat{R}_i^2$	$I - 1$	MS_R/MS_{res}
Column Main	$SS_C = I\sum_j \hat{C}_j^2$	$J - 1$	MS_C/MS_{res}
Row by Column	$SS_{RC} = \sum_i \sum_j y_{ij}^2 - SSm - SS_R - SS_C$	$(I - 1)(J - 1)$	
F_m	$SS_{F_m} = \hat{d}_m^2$	$I + J - 1 - 2m$	MS_{F_m}/MS_{res}
F_{res}	$SS_{F_{res}} = SS_{RC} - \sum_{m=1}^M SS_{F_m}$	$(I - 1 - M)(J - 1 - M)$	
Total	$SST = \sum_i \sum_j y_{ij}^2$	IJ	

A.2: Maximum likelihood estimation under the five models

This section shows derivation of maximum likelihood estimators (MLEs) for parameters in the four models listed previously. Define the parameter vector $\nu' = (\mu, R', C', \gamma', \sigma^2)$, the likelihood can then be expressed as

$$\begin{aligned}
 L(\nu) &= \left(\frac{1}{2\pi\sigma^2}\right)^{IJ/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_i \sum_j (y_{ij} - \mu - R_i - C_j - \gamma_{ij})^2\right\} \\
 &= \left(\frac{1}{2\pi\sigma^2}\right)^{IJ/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_i \sum_j [z_{ij} + (y_{..} - \mu) + (y_{i.} - y_{..} - R_i) + (y_{.j} - y_{..} - C_j) - \gamma_{ij}]^2\right\} \\
 &\quad \text{where, } (z_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}) \\
 &= \left(\frac{1}{2\pi\sigma^2}\right)^{IJ/2} \exp\left\{-\frac{1}{2\sigma^2} [IJ(y_{..} - \mu)^2 + J\sum_i (y_{i.} - y_{..} - R_i)^2 + I\sum_j (y_{.j} - y_{..} - C_j)^2 \right. \\
 &\quad \left. + \sum_i \sum_j (z_{ij} - \gamma_{ij})^2]\right\} \\
 &\leq \left(\frac{1}{2\pi\sigma^2}\right)^{IJ/2} \exp\left\{-\frac{1}{2\sigma^2} [\sum_i \sum_j (z_{ij} - \gamma_{ij})^2]\right\}
 \end{aligned}$$

The maximum value of $L(\nu)$ is attained when

$$\begin{aligned}
 \hat{\mu} &= y_{..} \\
 \hat{R}_i &= y_{i.} - y_{..}, i = 1, \dots, I \\
 \hat{C}_j &= y_{.j} - y_{..}, j = 1, \dots, J
 \end{aligned}$$

With the above estimates $\hat{\mu}, \hat{R}, \hat{C}$ substituted in the likelihood, the log likelihood that needs to be

maximized as a function of γ_{ij} is simply,

$$l^*(\gamma) = - \sum_i \sum_j \gamma_{ij}^2 + 2 \left(\sum_i \sum_j \gamma_{ij} z_{ij} \right). \quad (1)$$

Special Cases: Tukey 1-df: In this case, $\gamma_{ij} = \theta R_i C_j$ (Tukey 1949),

$$\begin{aligned} l^*(\theta) &= - \theta^2 \sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2 + 2\theta \sum_i \sum_j \hat{R}_i \hat{C}_j z_{ij} \\ \frac{\partial l^*(\theta)}{\partial \theta} &= - 2\theta \sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2 + 2 \sum_i \sum_j \hat{R}_i \hat{C}_j z_{ij} = 0 \\ \hat{\theta} &= \frac{\sum_i \sum_j \hat{R}_i \hat{C}_j z_{ij}}{\sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2} \end{aligned}$$

Mandel's Column Model: In this case, $\gamma_{ij} = \lambda_i C_j$ (Mandel 1961),

$$\begin{aligned} l^*(\lambda) &= - \sum_i \sum_j \lambda_i^2 \hat{C}_j^2 + 2 \sum_i \sum_j \lambda_i \hat{C}_j z_{ij} \\ \frac{\partial l^*(\lambda)}{\partial \lambda_i} &= - 2\lambda_i \sum_j \hat{C}_j^2 + 2 \sum_j \hat{C}_j z_{ij} = 0 \\ \hat{\lambda}_i &= \frac{\sum_j \hat{C}_j z_{ij}}{\sum_j \hat{C}_j^2} \end{aligned}$$

Principal Interactions or AMMI Model: In this case, $\gamma_{ij} = d_1 \alpha_i \beta_j$. Johnson and Graybill (1972) showed that the maximum likelihood for AMMI model with $M = 1$ is attained when the MLE \hat{d}_1^2 is equal to the largest root of $Z'Z$ (l_1). We present the following simpler argument to derive MLEs for the interaction terms.

Note that in that case, from (1),

$$l^*(d_1, \alpha_i, \beta_j) = -d_1^2 \sum_i \alpha_i^2 \sum_j \beta_j^2 + 2d_1 \sum_i \sum_j \alpha_i \beta_j z_{ij}.$$

Under the constraints, $\sum_i \alpha_i^2 = \sum_j \beta_j^2 = 1$, this reduces to maximizing $(\sum_i \sum_j \alpha_i \beta_j z_{ij})^2$ subject to the normalization constraints. It is well-known from the eigen-theory of matrices that this quadratic form is maximized by the values of $\{\alpha_i, \beta_j\}$ that are give by the left and right normalized characteristic vectors corresponding to $Z'Z$ and ZZ' where Z has the (i, j) -th entry $z_{ij} = \hat{\gamma}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}$. The maximum value of $(\sum_i \sum_j \alpha_i \beta_j z_{ij})^2$ is l_1 . With these estimates of α_i and β_j , it can be seen easily seen from the expression of $l^*(d_1, \alpha_i, \beta_j)$, that $\hat{d}_1^2 = l_1$. Hence the proof.

1 Description of Simulation Parameters

A Following are the row and column effects and other parameter setting for the simulation results in Sections 6.1 and 6.3 of the main text for the 3×3 Table scenario.

$$\begin{aligned}\mu &= 12; (R_1, R_2, R_3) = (C_1, C_2, C_3) = (-0.1, -0.2, 0.3), \\ \theta &= 1, (\lambda_1, \lambda_2, \lambda_3) = (\eta_1, \eta_2, \eta_3) = (-1, 0.8, 0.2), \quad d_1 = 1, \\ \alpha &= (-0.67, -0.07, 0.74), \\ \beta &= (-0.74, 0.07, 0.67).\end{aligned}$$

B: Following are the row and column effects and other parameter setting for the simulation results in Sections 6.1 of the main text for the 9×5 Table scenario. We highlight α_i and β_j in bold in the A and B matrices.

$$\begin{aligned}\mu &= 12 \\ R &= (-0.05, 0.04, 0.01, -0.10, 0.08, 0.02, 0.15, -0.12, -0.03) \\ C &= (0.55, 0.3, -0.05, -0.2, -0.6) \\ \theta &= 1.79 \\ \lambda &= \theta(0.4, 0.30, 1.92, -0.4, 0.5, -1.8, -0.53, -0.14, -0.25) \\ \eta &= \theta(0.1, -0.67, 0.49, 0.55, -0.47) \\ (d_1, d_2, d_3, d_4) &= (2.2, 0.8, 0.14, 0)\end{aligned}$$

$$A = \begin{pmatrix} \mathbf{0.23} & -0.97 & -0.07 & 0.00 \\ -\mathbf{0.09} & -0.02 & 0.02 & -0.40 \\ -\mathbf{0.37} & -0.10 & 0.09 & 0.86 \\ \mathbf{0.23} & 0.08 & -0.28 & 0.11 \\ -\mathbf{0.56} & -0.08 & -0.76 & -0.14 \\ -\mathbf{0.19} & -0.05 & 0.04 & -0.07 \\ \mathbf{0.37} & 0.10 & -0.09 & 0.14 \\ -\mathbf{0.19} & -0.05 & 0.04 & -0.07 \\ -\mathbf{0.47} & -0.15 & 0.56 & -0.22 \end{pmatrix}$$

$$B' = \begin{pmatrix} \mathbf{0.33} & \mathbf{0.87} & -\mathbf{0.36} & \mathbf{0.00} & \mathbf{0.00} \\ -0.50 & 0.00 & -0.44 & 0.73 & -0.15 \\ -0.25 & 0.00 & -0.22 & -0.47 & -0.82 \\ -0.58 & 0.49 & 0.65 & 0.00 & 0.00 \end{pmatrix}$$

A.4: Exact numerical tables for simulation results

Table 4: Percentage of interactions detected (or null hypotheses of no interaction rejected) by each of the five tests in the simulation settings corresponding to 3×3 array from 1000 simulated datasets with $N=1800$, 3600 and four repeated measures for each individual. Details are described in Section 6. The error variance σ_e^2 is set at 4 in all cases.

N	Simulation Model	σ_b^2	T-1DF	M-R	M-C	AMMI	Gollob
1800	Tukey 1-df	1	99.4	92.4	89.5	23.9	37.7
1800	Mandel-Row	1	0.0	87.4	0.0	17.7	28.2
1800	Mandel-Col	1	0.0	0.0	78.3	13.4	21.8
1800	AMMI (M=1)	1	0.0	0.0	0.0	18.7	30.0
1800	Additive*	1	18.1	14.2	10.9	6.3	10.2
1800	Tukey 1-df	4	90.9	68.8	63.7	16.8	25.6
1800	Mandel-Row	4	0.0	61.8	0.0	11.5	18.8
1800	Mandel-Col	4	0.0	0.0	49.6	10.0	15.3
1800	AMMI (M=1)	4	0.5	0.0	0.2	11.0	17.9
1800	Additive*	4	16.7	14.5	11.1	6.9	10.6
3600	Tukey 1-df	1	100.0	98.9	97.7	32.9	49.9
3600	Mandel-Row	1	0.0	98.3	0.0	24.5	38.6
3600	Mandel-Col	1	0.0	0.0	92.1	18.2	29.0
3600	AMMI (M=1)	1	0.0	0.0	0.0	25.4	39.7
3600	Additive*	1	18.7	14.7	10.9	5.3	8.4
3600	Tukey 1-df	4	99.4	88.6	83.9	21.3	31.6
3600	Mandel-Row	4	0.0	85.2	0.0	15.7	25.6
3600	Mandel-Col	4	0.0	0.0	70.5	11.2	20.2
3600	AMMI (M=1)	4	0.0	0.0	0.0	19.1	28.7
3600	Additive*	4	18.1	14.6	10.6	5.7	10.0

T-1DF = Tukey's 1 d.f. nonadditivity model; M-R = Mandel's row model;
M-C = Mandel's column model

*Type I error

Table 5: Percentage of interactions detected (or null hypotheses of no interaction rejected) by each of the five tests in the simulation settings corresponding to 9×5 array from 1000 simulated datasets with $N=3600, 7200$ and four repeated measures for each individual. Details are described in Section 6. The error variance σ_e^2 is set at 4 in all cases.

N	Simulation Model	σ_b^2	T-1DF	M-R	M-C	AMMI	Gollob
3600	Tukey-1DF	1	100.0	99.3	99.2	87.2	99.1
3600	Mandel-Row	1	0.3	99.6	0.2	86.2	99.2
3600	Mandel-Col	1	0.4	0.0	99.0	85.6	98.5
3600	AMMI	1	1.6	0.6	0.7	85.5	99.1
3600	Tukey-1DF	4	94.1	85.6	78.4	47.5	87.2
3600	Mandel-Row	4	3.6	84.5	1.2	45.2	86.6
3600	Mandel-Col	4	3.1	1.4	78.0	37.5	85.0
3600	AMMI	4	3.4	2.8	4.0	46.1	87.1
7200	Tukey-1DF	1	100.0	100.0	100.0	99.3	100.0
7200	Mandel-Row	1	0.1	100.0	0.0	99.4	100.0
7200	Mandel-Col	1	0.0	0.0	100.0	99.3	100.0
7200	AMMI	1	0.1	0.0	0.0	99.6	100.0
7200	Tukey-1DF	4	99.6	98.8	98.2	77.7	97.9
7200	Mandel-Row	4	0.9	99.3	0.1	78.9	98.1
7200	Mandel-Col	4	0.9	0.0	98.1	74.3	98.1
7200	AMMI	4	0.8	1.2	1.5	78.7	98.7

T-1DF = Tukey's 1 d.f. nonadditivity model; M-R = Mandel's row model;
M-C = Mandel's column model

Table 6: Percentage of interactions detected (or null hypotheses of no interaction rejected) by each of the five tests in 1000 simulated datasets under 10 common epistasis models. The error variance σ_e^2 is set at 4 in all cases. N=1800, 3600.

N	Epistasis Model	$\sigma_b^2 = 1$					$\sigma_b^2 = 4$				
		T-IDF	M-R	M-C	AMMI	Gollob	T-IDF	M-R	M-C	AMMI	Gollob
1800	Dom or Dom	100.0	94.5	95.4	33.0	51.5	96.4	75.1	74.2	22.6	35.1
1800	Dom or Rec	99.8	98.2	88.0	25.0	39.2	94.8	81.2	59.7	12.9	21.3
1800	Modified	97.0	98.5	64.8	25.6	39.9	81.1	84.6	39.1	16.8	25.9
1800	Dom and Dom	99.8	93.6	83.4	35.8	51.0	90.2	72.3	54.0	20.9	35.0
1800	Rec or Rec	99.8	77.6	87.5	12.0	20.6	91.8	45.4	59.3	10.0	15.4
1800	Dom and Rec	97.0	87.0	67.7	22.7	36.5	77.4	61.5	42.8	19.1	28.0
1800	Rec and Rec	87.8	57.9	58.1	16.6	23.4	56.0	29.1	32.8	8.4	13.9
1800	Checkboard	98.2	79.8	85.5	28.4	43.7	83.9	56.0	61.4	18.8	28.9
1800	Add and Add	10.0	17.3	17.0	63.2	84.7	10.9	14.0	16.0	40.6	62.8
1800	General	14.5	24.1	17.4	89.6	98.9	16.0	24.3	18.5	69.2	89.6
3600	Dom or Dom	100.0	99.2	99.7	46.6	66.5	100.0	92.2	93.5	30.2	47.9
3600	Dom or Rec	100.0	99.9	97.4	35.5	54.8	99.2	96.8	81.4	21.8	35.4
3600	Modified	100.0	100.0	83.8	34.4	51.8	95.2	98.0	60.7	24.0	34.5
3600	Dom and Dom	100.0	99.3	94.2	46.0	66.0	99.5	90.9	75.2	28.5	46.7
3600	Rec or Rec	100.0	94.1	96.7	18.4	32.2	99.4	69.1	82.6	13.1	21.7
3600	Dom and Rec	99.7	97.4	83.7	34.2	51.3	94.5	81.2	61.7	23.4	36.7
3600	Rec and Rec	97.6	77.8	77.9	20.4	32.8	81.3	50.2	50.8	13.9	21.4
3600	Checkboard	99.8	93.9	95.7	37.1	56.5	97.3	76.1	77.4	25.1	39.4
3600	Add and Add	10.6	15.2	18.1	78.7	95.2	12.4	16.1	17.0	56.2	79.0
3600	General	16.8	25.5	16.7	97.7	100.0	16.8	25.6	16.8	85.7	98.2

T-IDF = Tukey's 1 d.f. nonadditivity model; M-R = Mandel's row model; M-C = Mandel's column model

A.3 Supplementary Figure

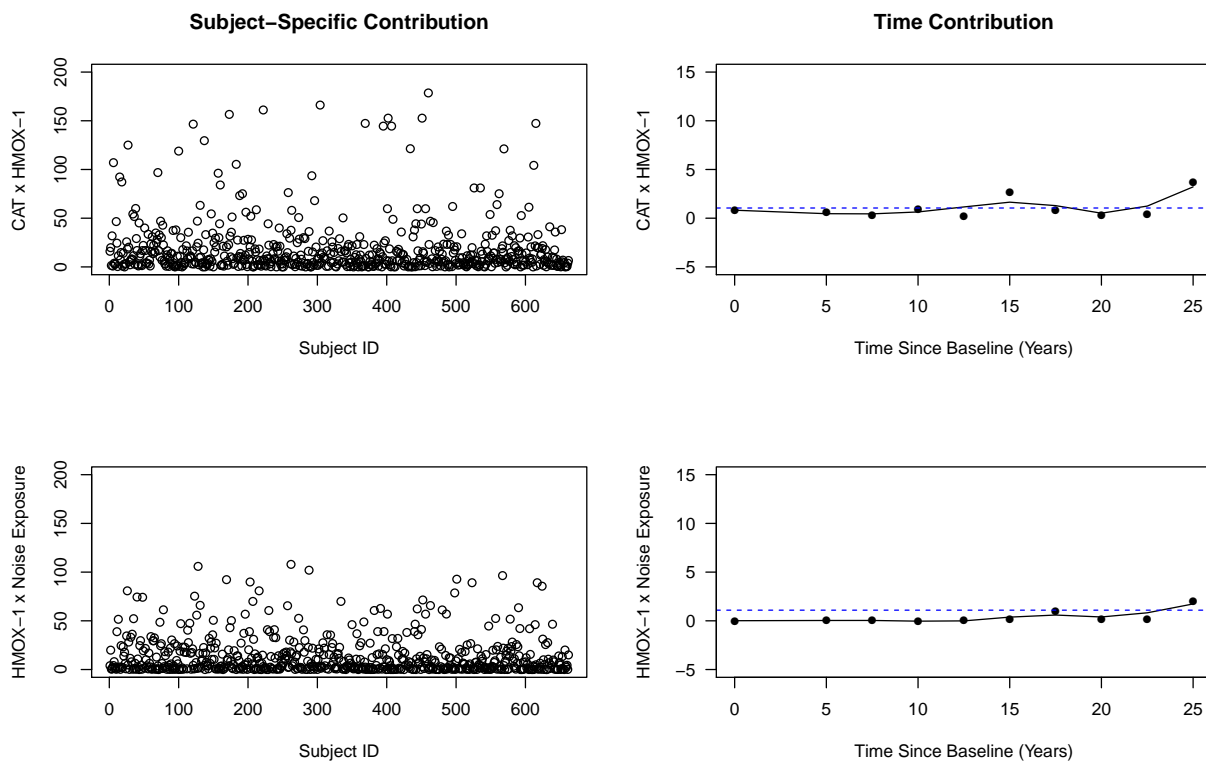


Figure 1: Subject-specific contribution and time contribution to the second interaction factor in gene-gene interaction (upper panel) and gene-environment interaction (lower panel) in the Normative Aging Study.