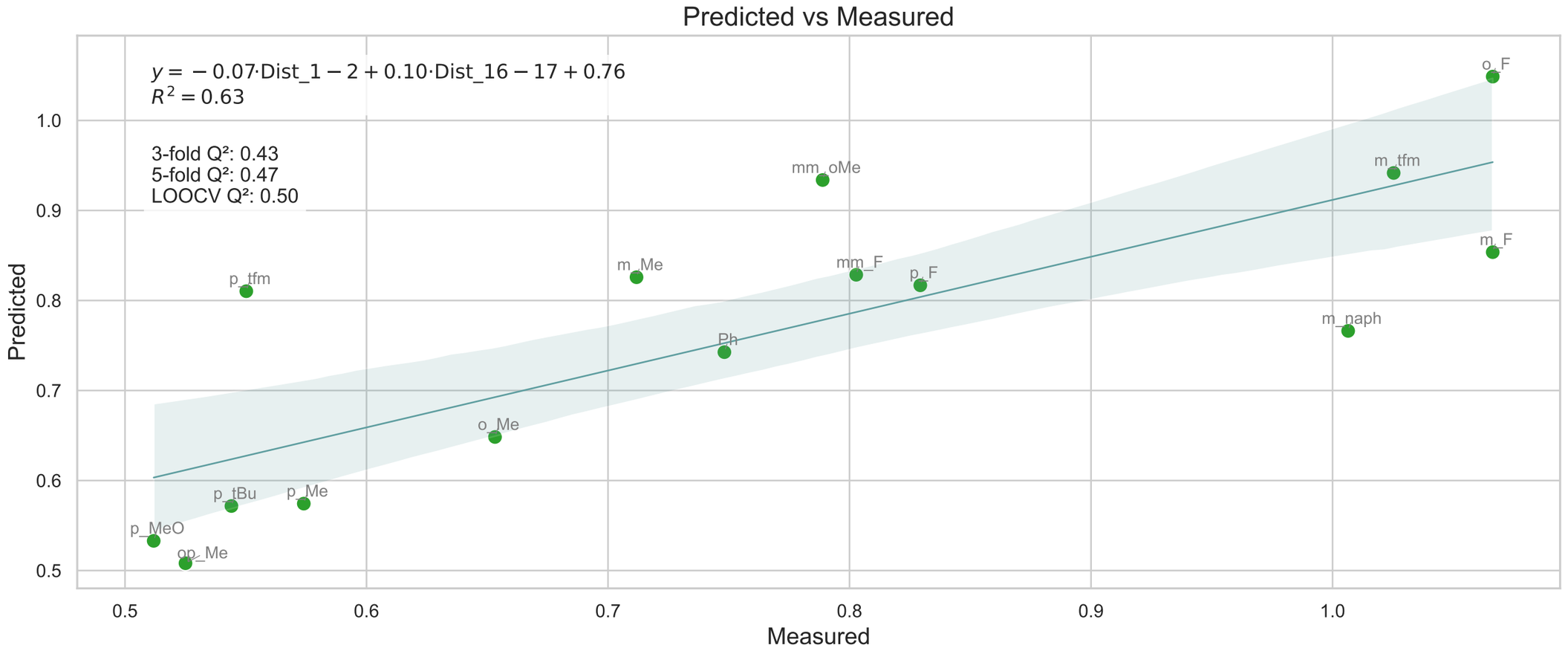

Predicted vs Measured
$y = -0.07 \cdot \text{Dist_1} - 2 + 0.10 \cdot \text{Dist_16} - 17 + 0.76$
$R^2 = 0.63$
3-fold Q²: 0.43
5-fold Q²: 0.47
LOOCV Q²: 0.50
Predicted
Measured
0.5
0.6
0.7
0.8
0.9
1.0
o_F
m_tfm
mm_oMe
m_F
m_Me
mm_F
p_F
p_tfm
m_naph
Ph
o_Me
p_tBu
p_Me
p_MeO
op_Me

3-fold CV Metrics

Q2_3_Fold	MAE_3	RMSD_3
0.434	0.103	0.147

5-fold CV Metrics

Q2_5_Fold	MAE_5	RMSD_5
0.474	0.099	0.142

LOOCV Metrics

Q2_LOOCV	MAE_LOOCV	RMSD_LOOCV
0.496	0.095	0.14

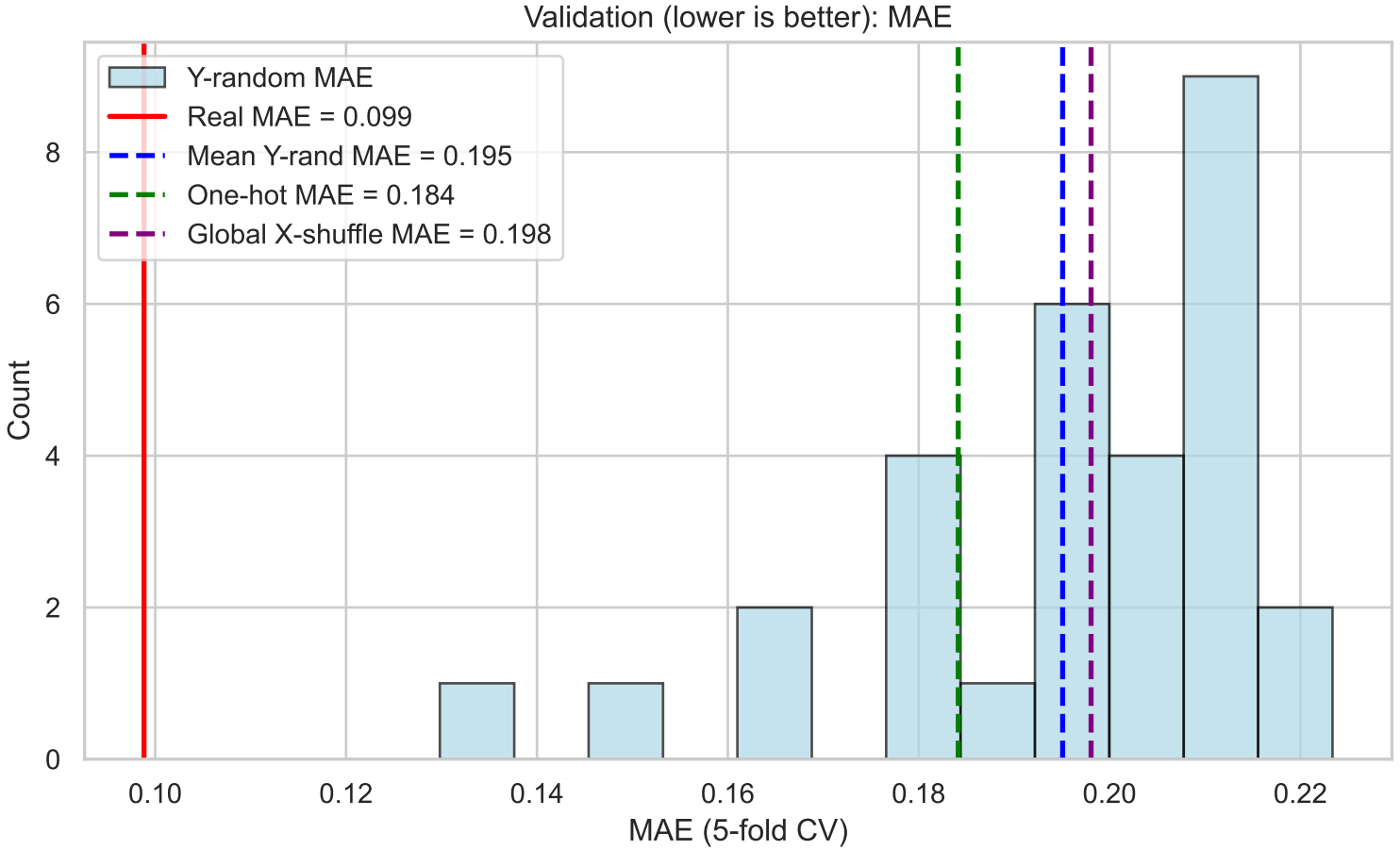
Variance Inflation Factors

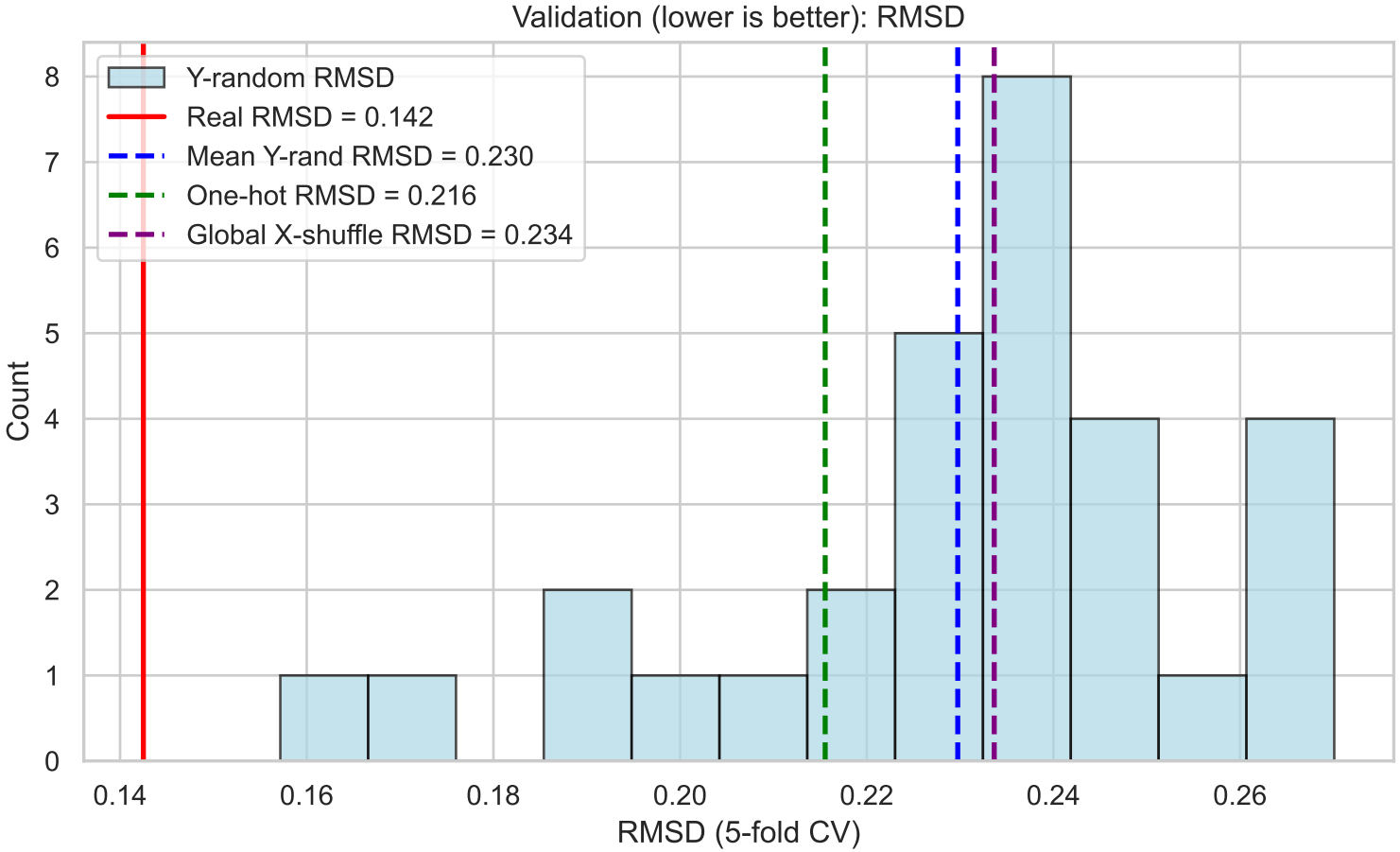
variable	VIF
Dist_1-2	1.69
Dist_16-17	1.69

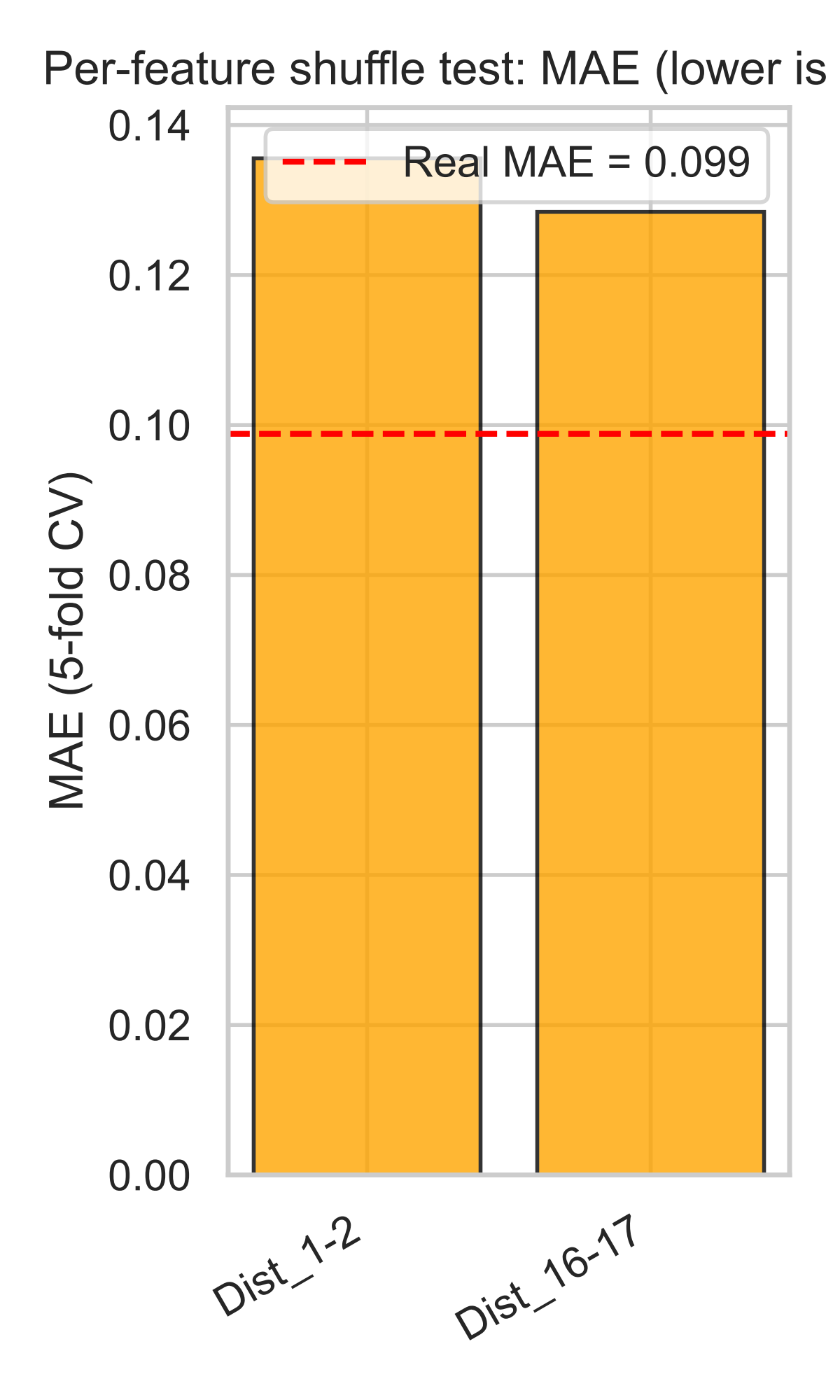
Coefficient Estimates

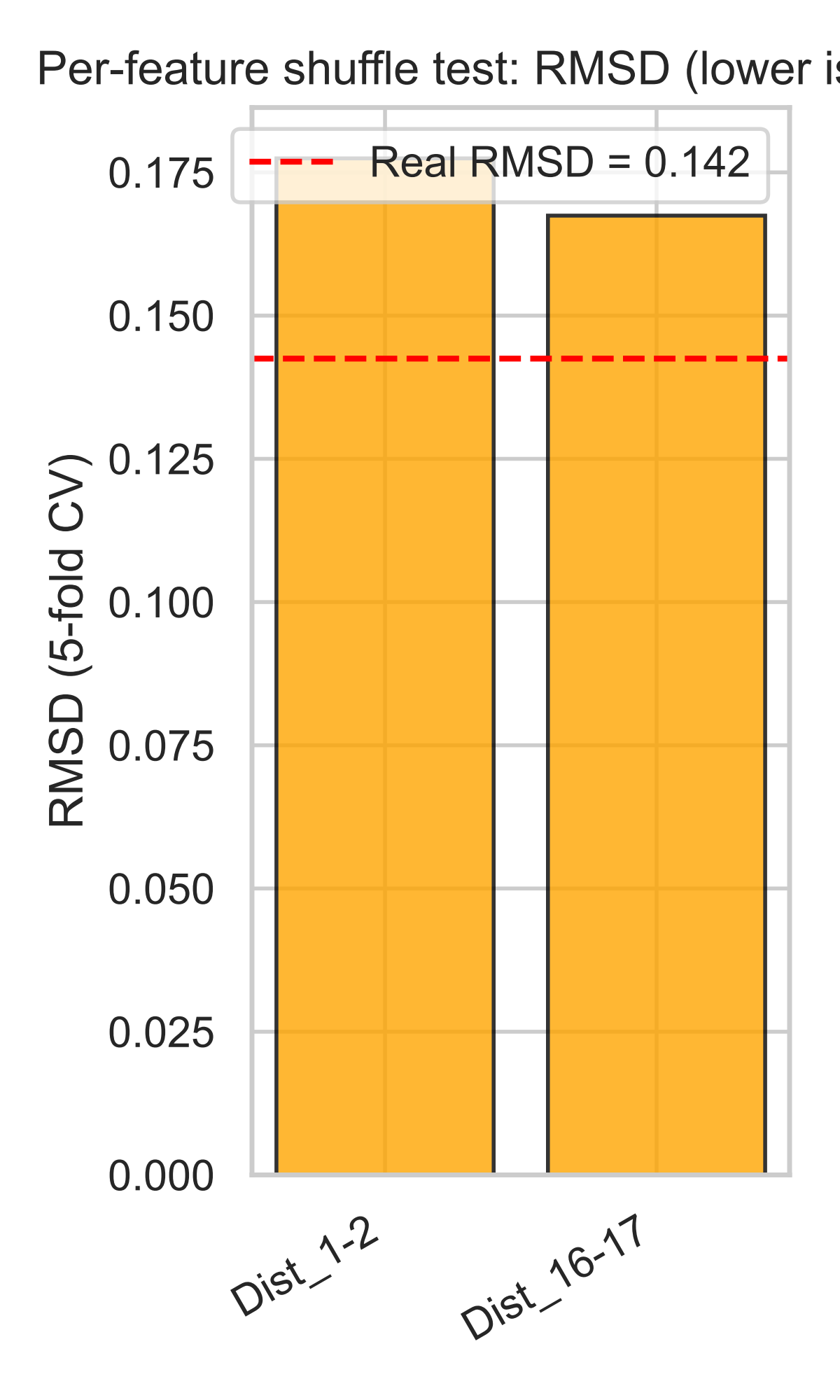
Estimate	Std. Error	t value	p value
0.76	0.035	22.035	0.0
-0.073	0.045	-1.621	0.131
0.1	0.045	2.227	0.046

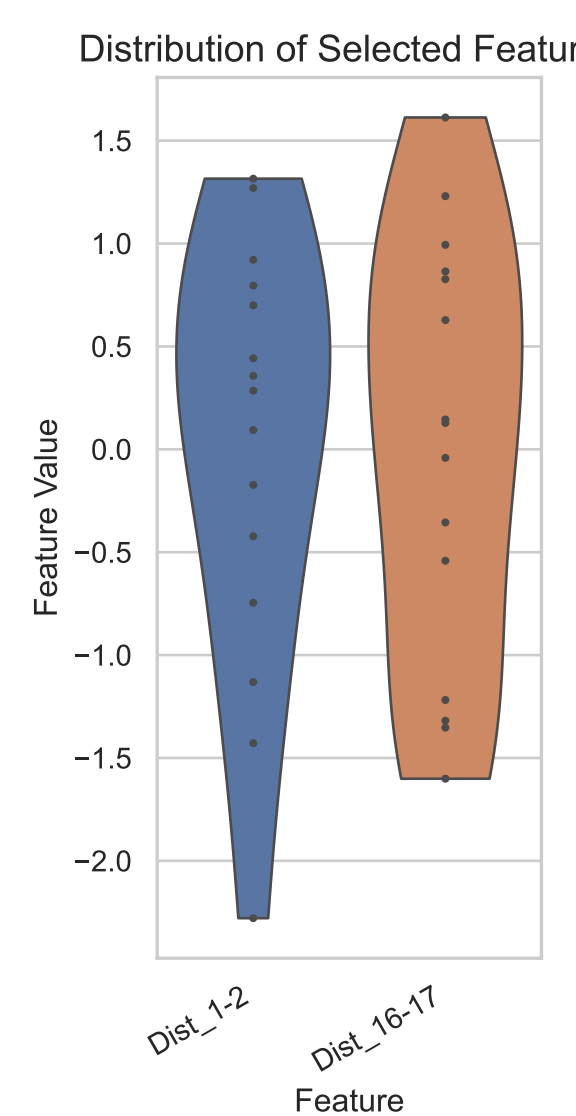
Left-out Predictions (none)

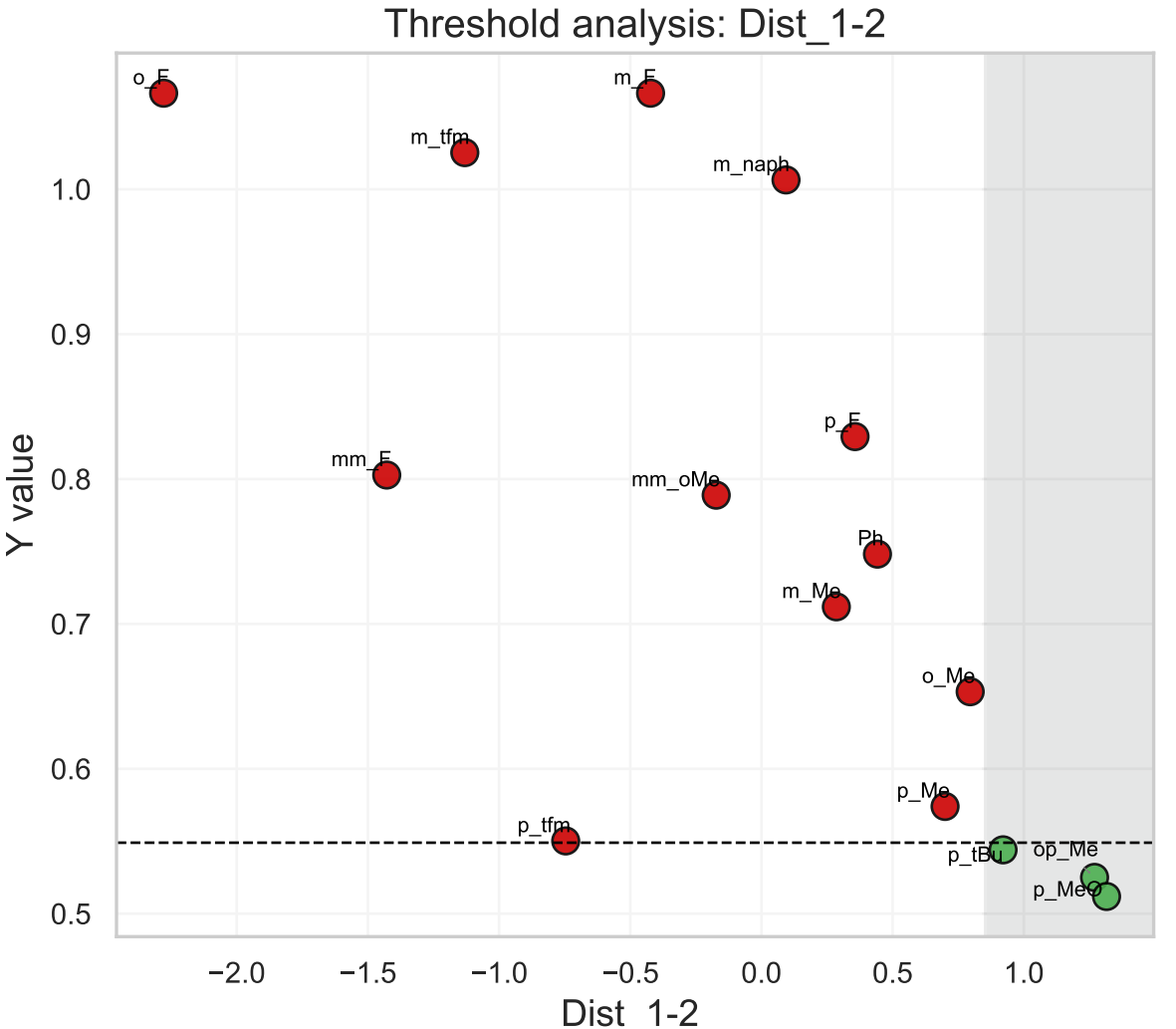

Validation (lower is better): MAE
Y-random MAE
Real MAE = 0.099
Mean Y-rand MAE = 0.195
One-hot MAE = 0.184
Global X-shuffle MAE = 0.198
Count
0
2
4
6
8
0.10
0.12
0.14
0.16
0.18
0.20
0.22
MAE (5-fold CV)

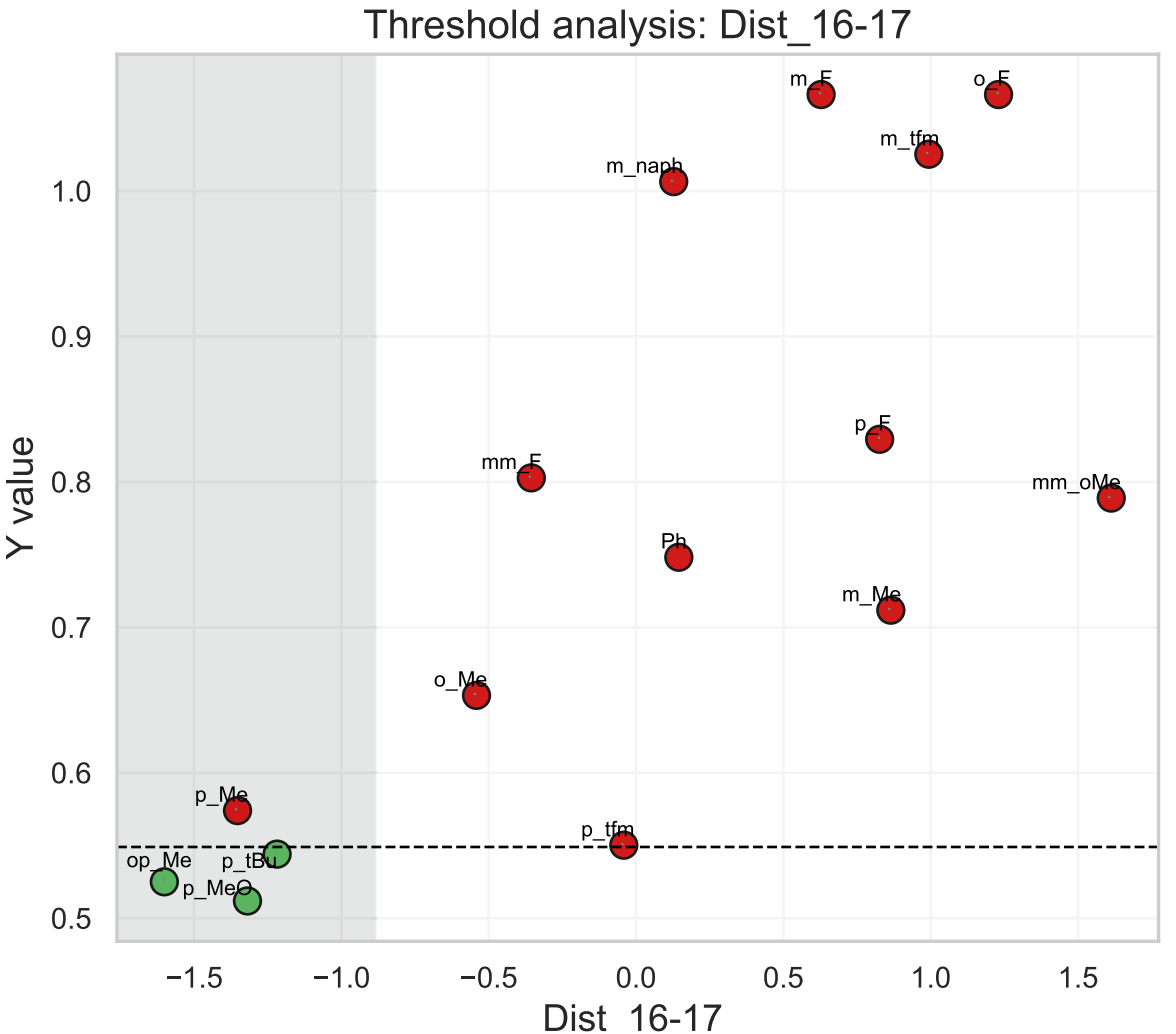

Validation (lower is better): RMSD
Y-random RMSD
Real RMSD = 0.142
Mean Y-rand RMSD = 0.230
One-hot RMSD = 0.216
Global X-shuffle RMSD = 0.234
Count
RMSD (5-fold CV)
0
1
2
3
4
5
6
7
8
0.14
0.16
0.18
0.20
0.22
0.24
0.26


Per-feature shuffle test: MAE (lower is
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
MAE (5-fold CV)
Real MAE = 0.099
Dist_1-2
Dist_16-17


Per-feature shuffle test: RMSD (lower is
Real RMSD = 0.142
RMSD (5-fold CV)
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Dist_1-2
Dist_16-17


Distribution of Selected Featur
Feature Value
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
Dist_1-2
Dist_16-17
Feature


Threshold analysis: Dist_1-2
o_F
m_F
m_tfm
m_naph
p_F
mm_F
mm_oMe
Ph
m_Me
o_Me
p_Me
p_tfm
p_tBu
op_Me
p_MeO
Y value
1.0
0.9
0.8
0.7
0.6
0.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
Dist_1-2


Threshold analysis: Dist_16-17
m_F
o_F
m_tfm
m_naph
p_F
mm_F
mm_oMe
Ph
m_Me
o_Me
p_Me
p_tfm
op_Me
p_tBu
p_MeO
Y value
1.0
0.9
0.8
0.7
0.6
0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Dist_16-17