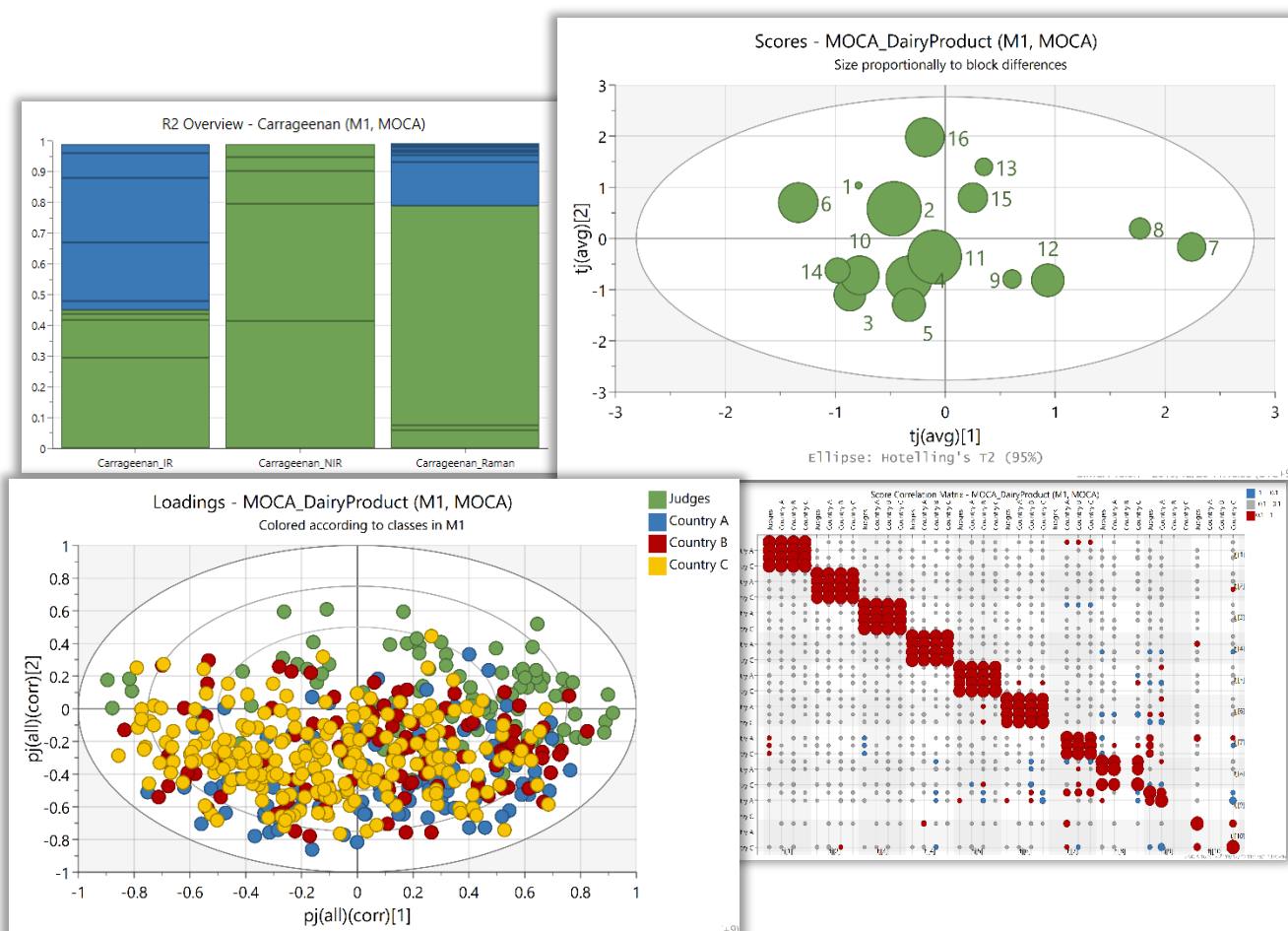


SIMCA17 MOCA チュートリアル

～食品評価データ編～



infocom

インフォコム株式会社
ライフサイエンスグループ

目次

0. はじめに	1
1. データセット.....	3
2. MOCA モデルを作成する	6
モデルの設定	6
モデルの計算結果の解釈	11
モデルの表示	13
3. 結果を解釈する	20
4. 最後に	25
(参考) 既知の問題	25

0. はじめに

MOCA とは

MOCA は **M**ultiblock **O**rthogonal **C**omponent **A**nalysis の略で、複数のデータ群（マルチブロックデータ）を解析する新しい手法です。

MOCA では、全てのブロックでサンプルが共通し、変数のみが異なるマルチブロックデータの解析が可能です。

OnPLS, JUMBA のアルゴリズムを基に、共通の変動を取り出すことで、データの概観・特徴を捉えることを可能にしています。

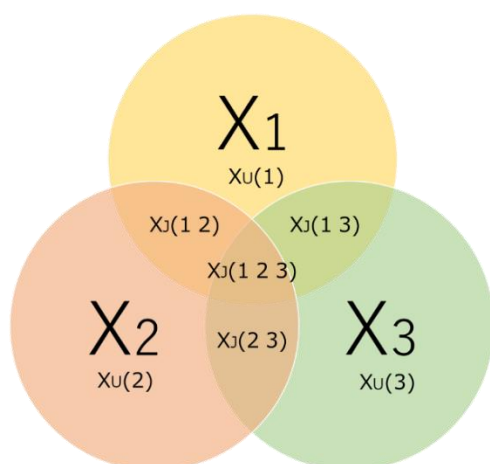
マルチブロックデータの解析例：

- ・ マルチオミクス： メタボロミクスデータ ・ プロテオミクスデータ ・ リピドミクスデータ
- ・ 製造データ： プロセス信号 ・ 分光法 ・ 原料組成データ
- ・ 製品評価： 官能評価 ・ 化学分析 ・ 消費者の好み

MOCA は、SIMCA15 以前の O2PLS の拡張版であり、これまでは 2 ブロックデータまでの解析が可能でしたが、MOCA では 3 ブロック以上での解析も可能となりました。

これまでの O2PLS は MOCA へ統合されました。

MOCA では各ブロック X_i について、下式のように分解します。（3 ブロックの例）



$$X_i = \underbrace{X_G}_{\text{Global Joint}} + \underbrace{X_L}_{\text{Local Joint}} + \underbrace{X_U}_{\text{Unique}} + \underbrace{E}_{\text{Residual}}$$

- ・ Globally Joint = 全ブロック共通の変動
- ・ Locally Joint = 一部共通の変動
- ・ Unique = 固有の変動
- ・ Residual = 残差

この原理に沿って、ブロックの共通変動及び固有変動を抽出し、包括的かつ信頼性の高い、マルチブロックデータ解析を可能にしています。

データ

以下のデータを用いて、官能評価及び消費者評価のマルチブロック直交成分分析（MOCA）を行います。



- ・ サンプル：機能性食品（ヨーグルト）の16製品サンプル
- ・ 4ブロックの各データ変数：
 - 官能評価 102項目（食感、香り、色、感覚のパネルデータ）
 - A国 消費者評価 113人（好き～嫌いを1～9点で評価）
 - B国 消費者評価 113人（〃）
 - C国 消費者評価 228人（〃）

これらのデータを用いて、消費者に好まれる製品を見つけることを目的とします。

1. データセット

データセットの準備

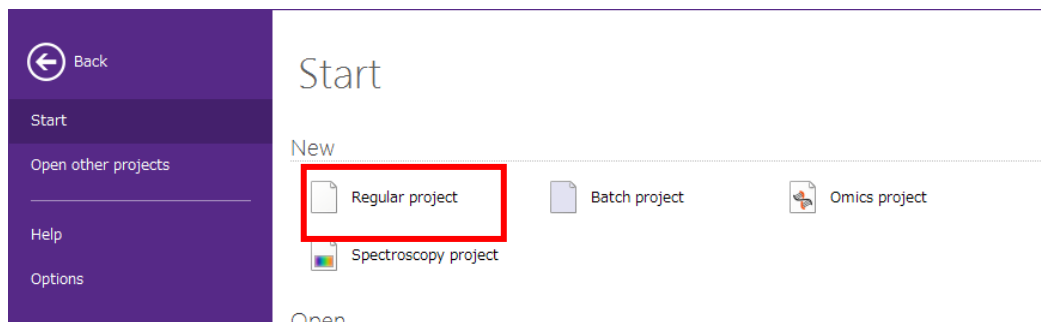
Excel 形式で用意します。

ファイル名：「MOCA_DairyProduct.xlsx」

Primary ID	Obs. Sec. ID	Aroma 1 Judges	Aroma 2 Judges	Aroma 3 Judges	Aroma 4 Judges	Aroma 5 Judges	Aroma 6 Judges	Aroma 7 Judges	Aroma 8 Judges	Aroma 9 Judges	Aroma 10 Judges	Aroma 11 Judges	Aroma 12 Judges	Aroma 13 Judges	Aroma 14 Judges
1	Product 1	0.00402482	0	0.0241489	0	0.474929005	0	0.0402482	0	0.088546097	0.00402482	0.112695001	0.0241489	0.00402482	0.0201241
2	Product 2	0.00459514	0	0.041356299	0	0.0137854	0	0.036761101	0.00229757	0.0919028	0.00229757	0	0.00919028	0.00229757	0.00229757
3	Product 3	0.0037397	0	0.0448764	0.00186985	0.0243081	0	0.0251779	0.00560956	0.069184497	0.0037397	0.00560956	0.013089	0.00186985	0.00186985
4	Product 4	0	0	0.032274898	0	0.107583001	0	0.048412301	0.00537915	0.064549796	0	0.0107583	0.016137401	0.00537915	0.00537915
5	Product 5	0.00319718	0	0.0287746	0	0.019183099	0	0.035169002	0.00319718	0.067140803	0.00319718	0.00319718	0.015985901	0.00319718	0
6	Product 6	0.00608579	0	0.048686299	0	0.0182574	0	0.00608579	0.0030429	0.030429	0.0030429	0.0030429	0.00608579	0.00608579	0.0030429
7	Product 7	0.00302667	0.00605333	0.0605333	0	0.00908	0.024213299	0.115013003	0.0423733	0.105933003	0.00605333	0.00908	0.172519997	0.0211867	0.00302667
8	Product 8	0.00594241	0.0029712	0.0564529	0	0.00891361	0	0.115877002	0.00594241	0.041596901	0.0029712	0.038625699	0.047539301	0.00594241	0.00594241
9	Product 9	0.0106573	0.0213146	0.063943699	0.037300501	0.00532864	0	0.111901	0.00532864	0.085258201	0.00532864	0.0266432	0.0266432	0	0.00532864
10	Product 10	0.00718354	0	0.0431013	0.00359177	0.050284799	0	0.082610801	0.00359177	0.082610801	0.00359177	0.147263005	0.061060101	0	0
11	Product 11	0.00548309	0	0.043864701	0	0.0137077	0	0.046806299	0.00548309	0.112402998	0.00274155	0.00548309	0.00622464	0.00274155	0.00274155
12	Product 12	0.00192047	0	0.0460914	0	0.00576142	0.00576142	0.044170901	0.0134433	0.132513002	0.00192047	0.00192047	0.098864602	0.0115228	0.00192047
13	Product 13	0.00439072	0	0.234902993	0.00219536	0.0285397	0.00878143	0.058059099	0.00219536	0.333694011	0.00219536	0.0131721	0.120745003	0.00439072	0.00439072
14	Product 14	0.00418819	0	0.058634698	0.00418819	0.0167528	0.00418819	0.071199298	0.00418819	0.0544465	0.00418819	0.196844995	0.025129201	0.033505499	0.00837638
15	Product 15	0.00281289	0	0.0281289	0	0.033754598	0.00281289	0.0267224	0.0098451	0.067509301	0.00281289	0.0140644	0.084386602	0.0140644	0.00140644
16	Product 16	0	0	0.061021499	0	0.299005002	0.00610215	0.103736997	0	0.042715099	0	0.805483997	0.152554005	0.00610215	0

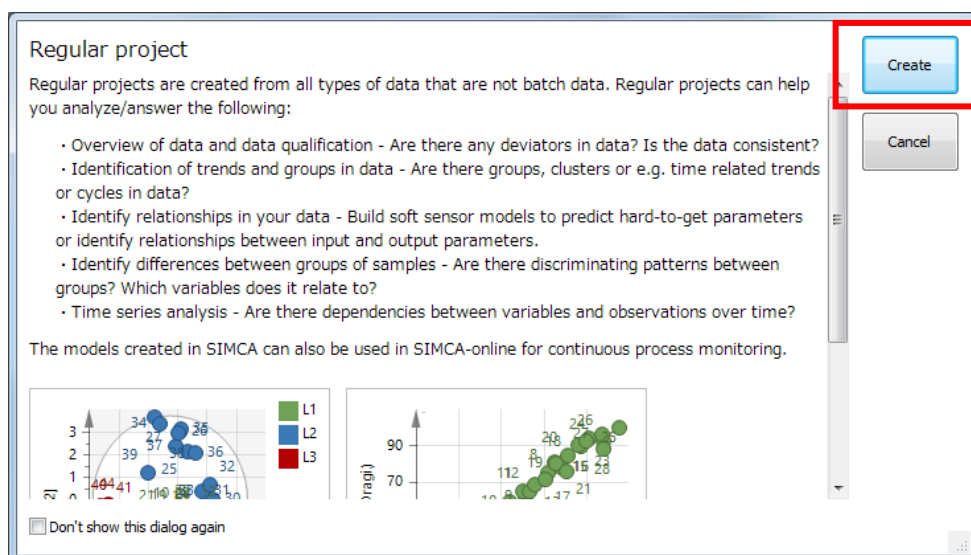
データセットのインポート

SIMCA 17 を起動します。



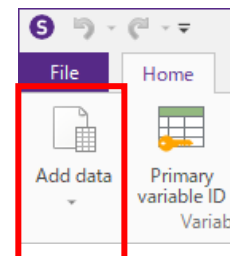
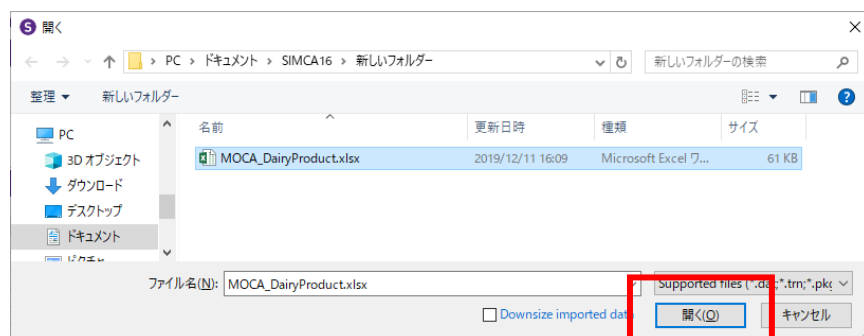
Start 画面が立ち上がるので、Regular project をクリックします。

Create をクリックします。



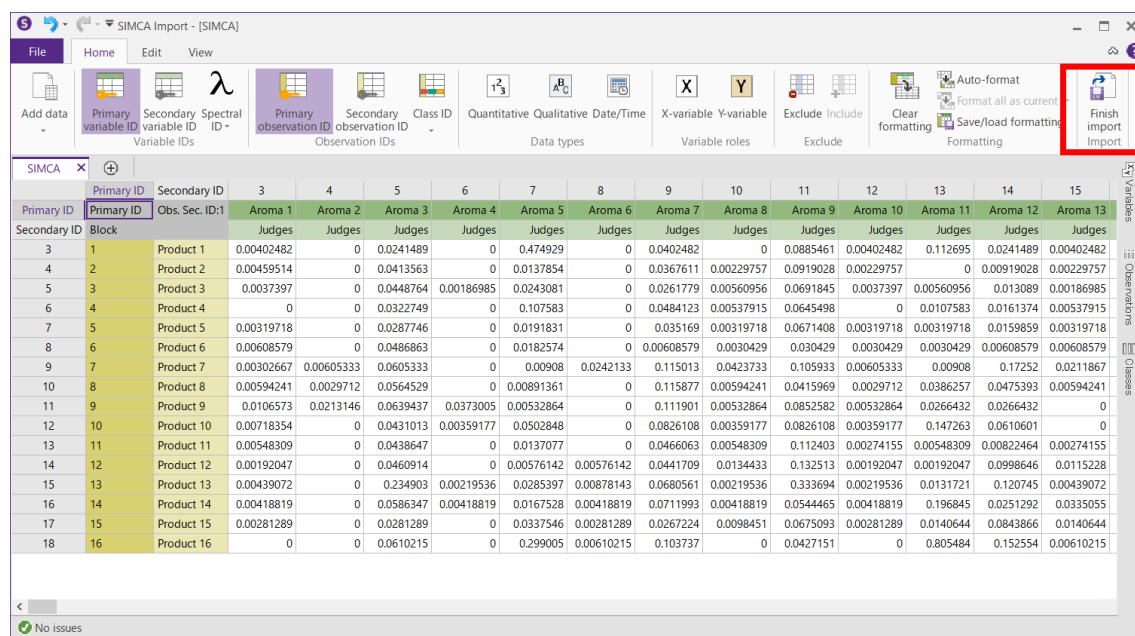
用意していたデータセットを指定して、開きます。

(このダイアログが開かない場合は、Home > Add data をクリックします。)



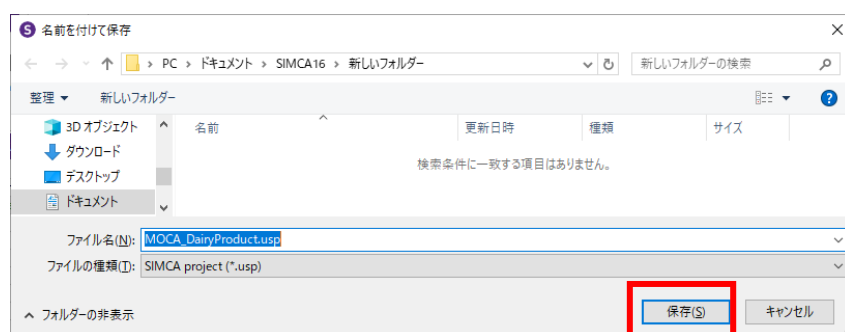
SIMCA が自動で ID を設定します。ここでは特に設定を変更する必要はありません。

データセットの設定が完了したので“Finish import”をクリック、保存先を選択し、インポートを終了します。



※設定時、Class ID は使用しないでください。

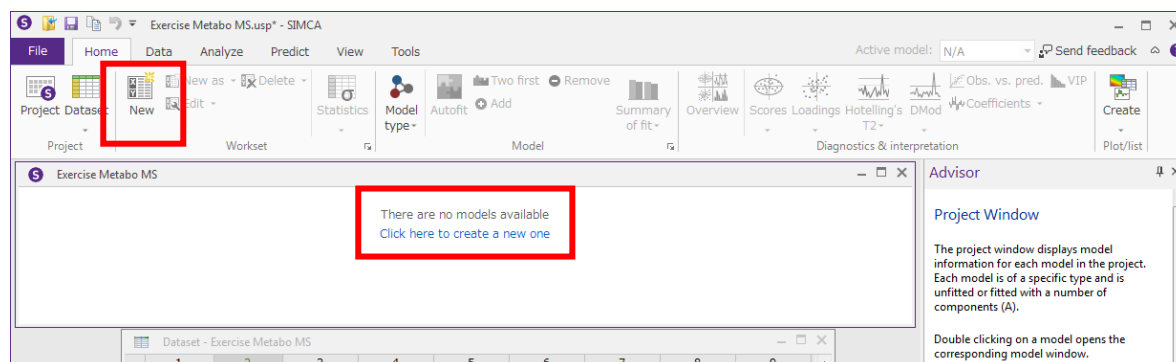
モデル作成時に、Block の設定ができなくなります。(詳細は P.25～)



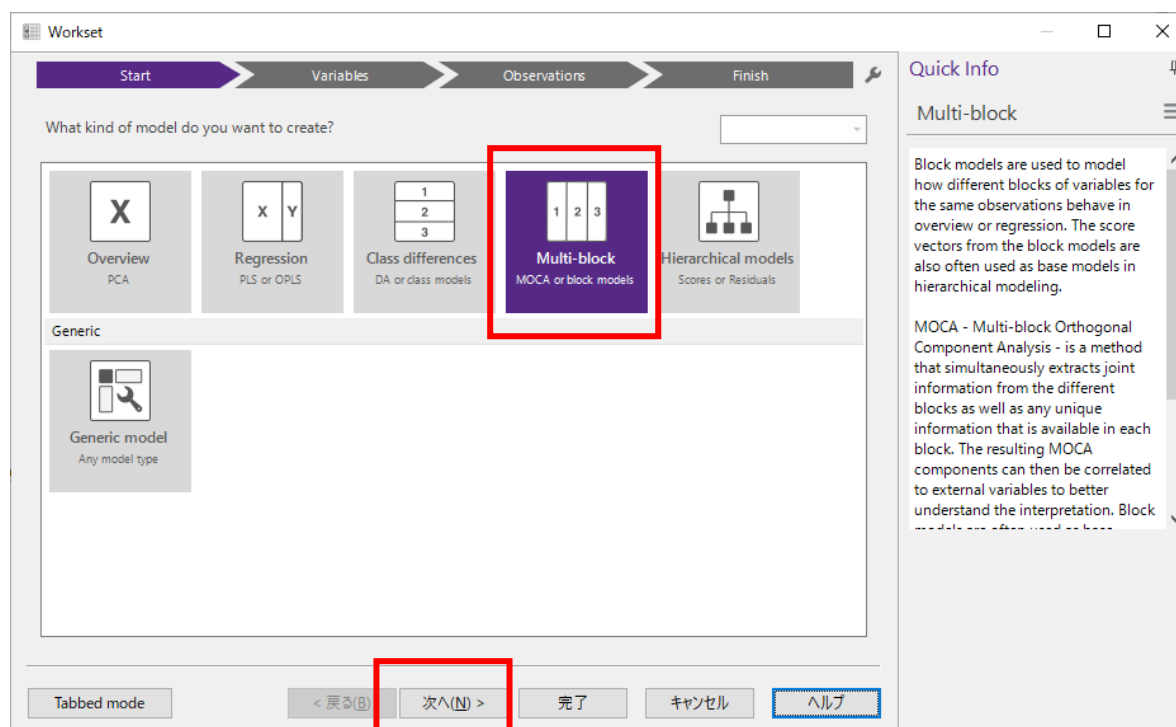
2. MOCA モデルを作成する

モデルの設定

Home > New または “Click here to create a new one” をクリックします。

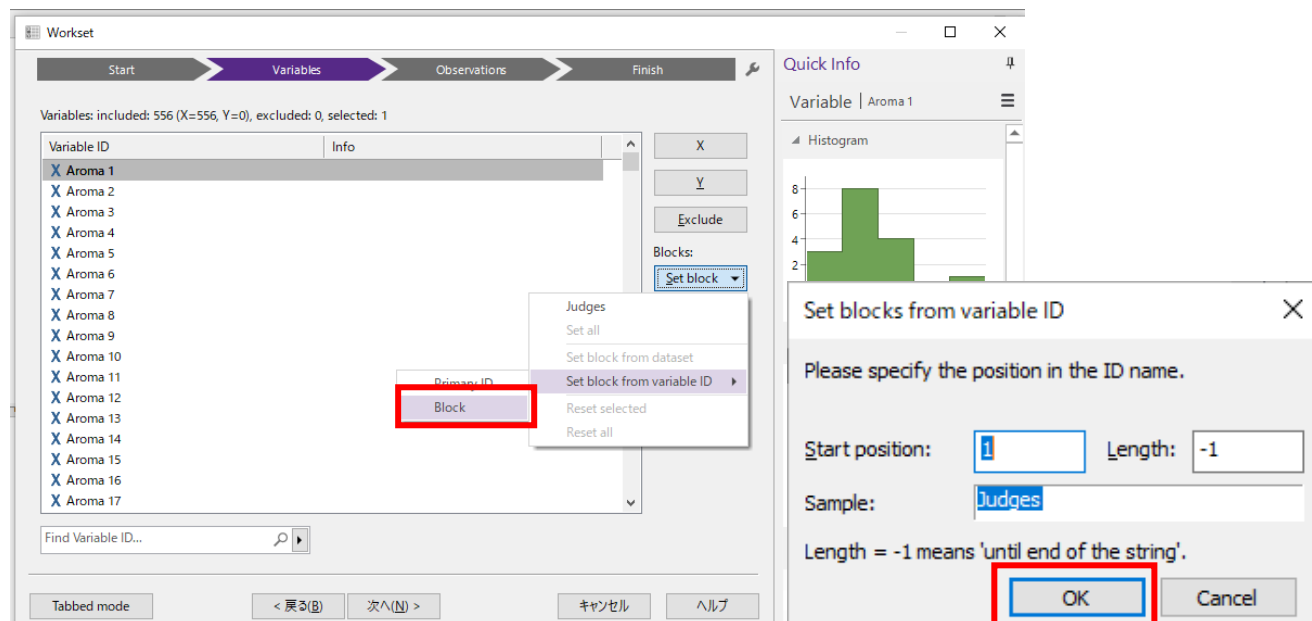


Multi-block を選択し、次へをクリックします。

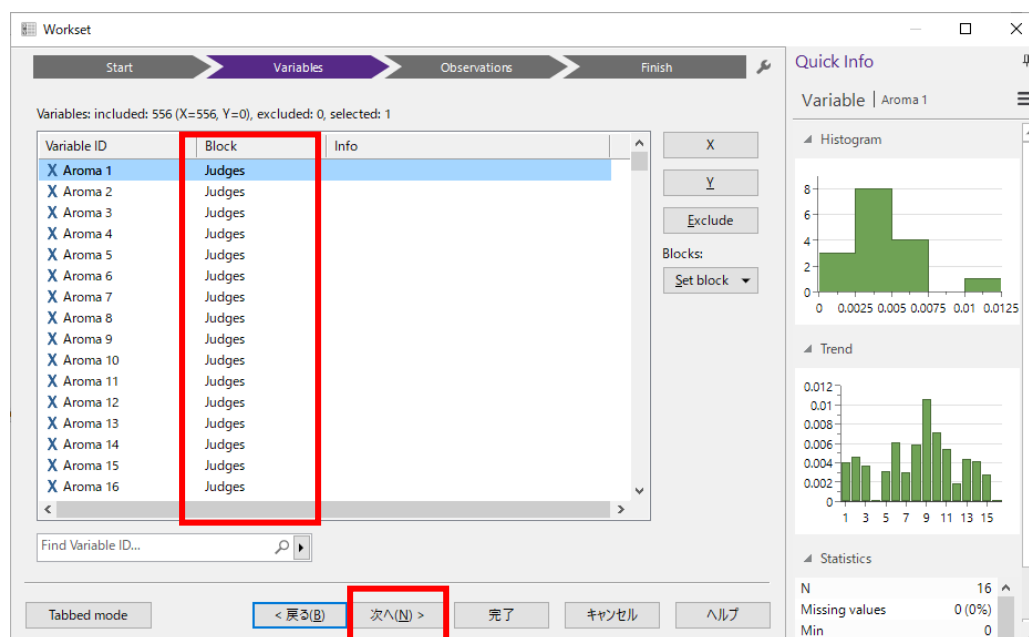


Variables（変数）の設定をします。変数はデフォルトのように、全て X で設定します。

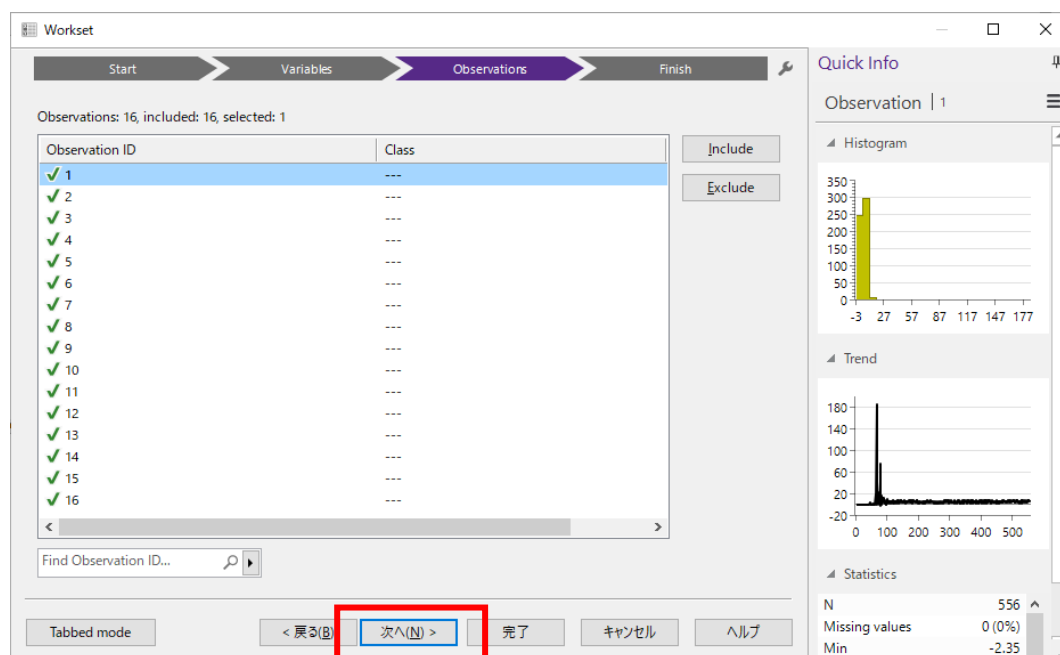
Set block > Set block from variable ID から “Block” を選択し、OK をクリックします。



Block が設定されたことを確認し、次へをクリックします。

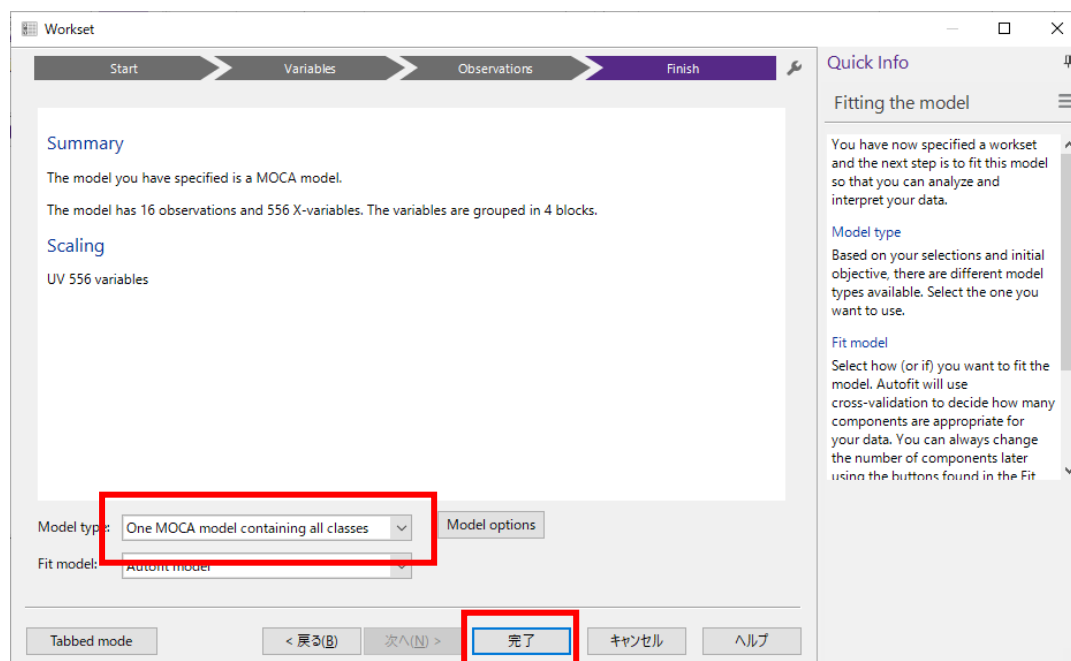


Observation（サンプル）の設定をします。ここでは特に設定の必要はありません。次へをクリックします。

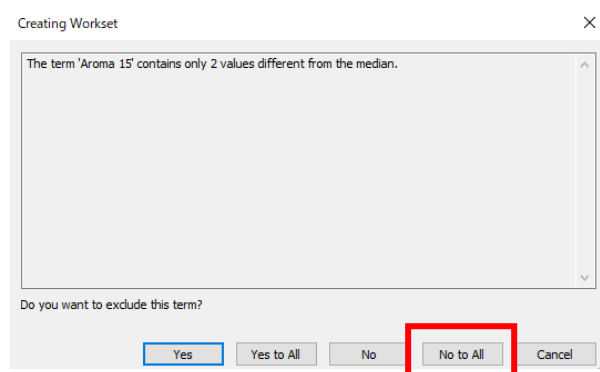


Summary 画面ではモデルタイプ、変数（variables）、サンプル（observations）、ブロック数およびスケーリング（scaling）が表示されます。MOCA ではデフォルトとして、UV スケーリングの使用が推奨されています。

Model type が “One MOCA model containing all classes” に設定されていることを確認し、完了をクリックします。

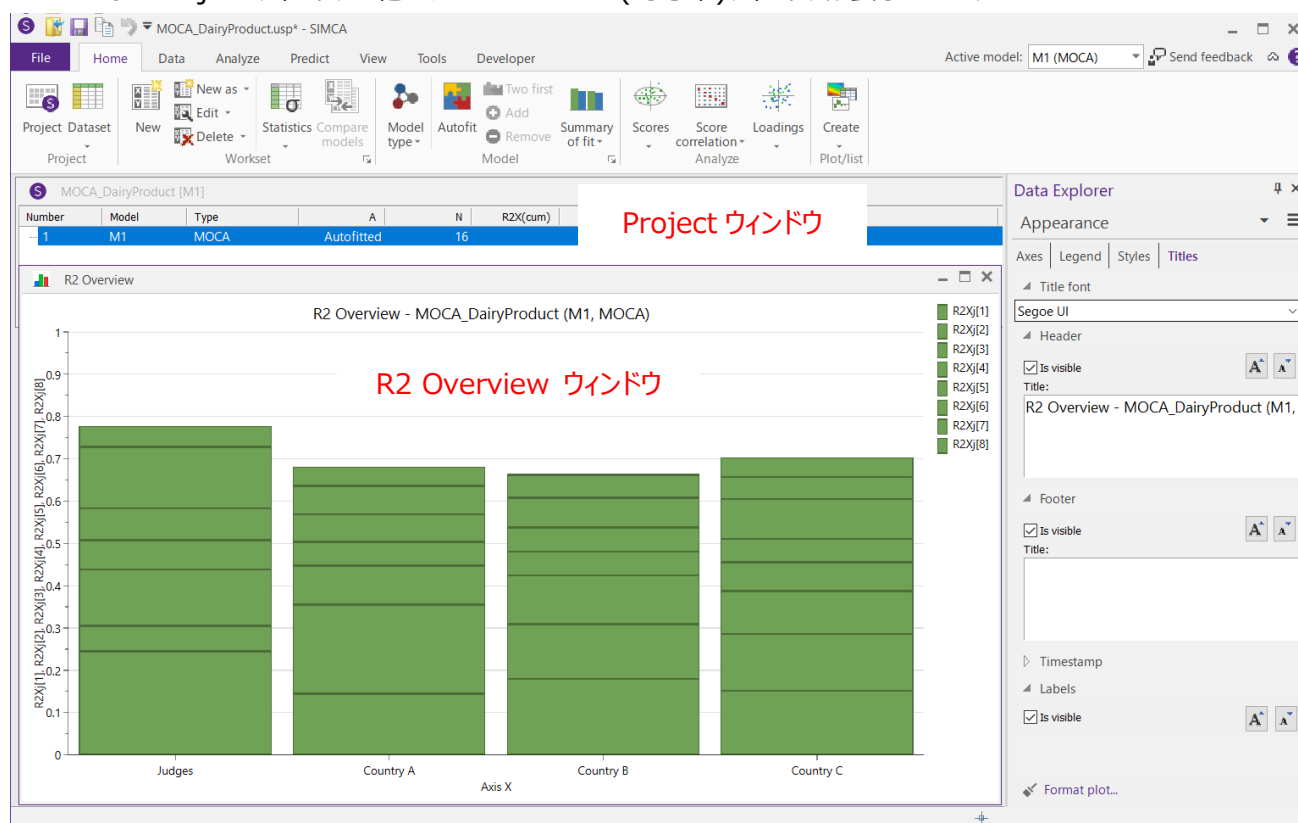


No to All をクリックします。



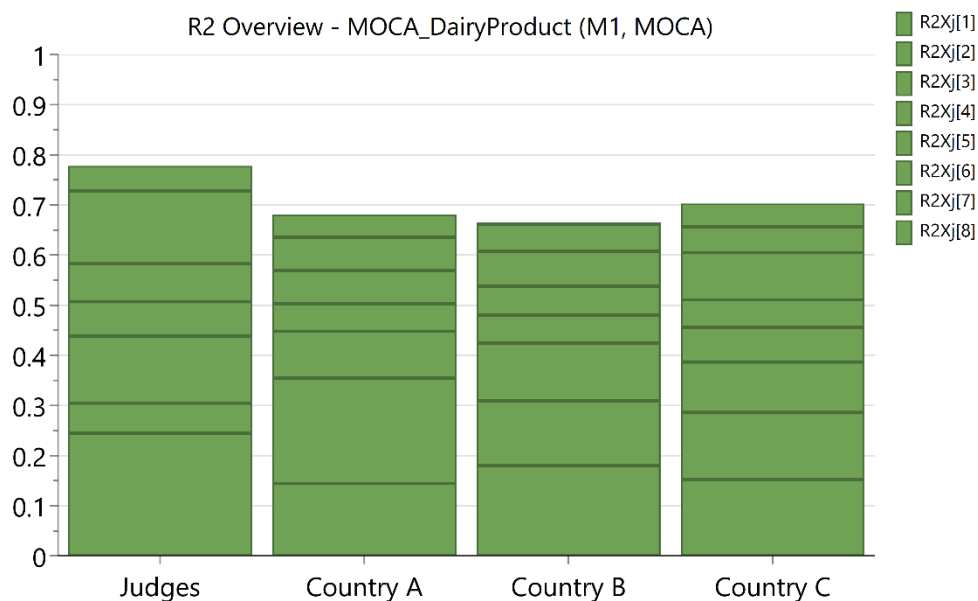
SIMCA が自動的に最適なモデルを計算し、モデルを作成します。

MOCA では Project ウィンドウの他に、R2 Overview(寄与率)ウィンドウが表示されます。

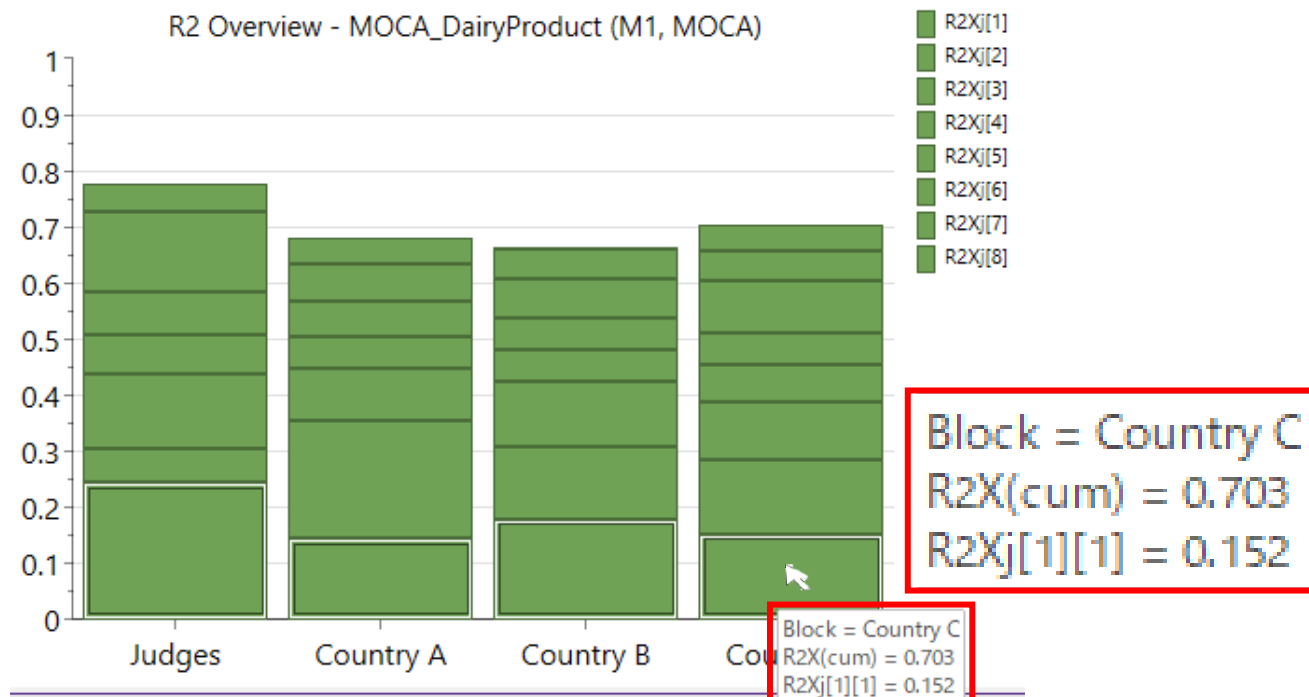


R2 Overview では、横軸は各ブロックを、縦軸は R2（寄与率）を示し、Joint（緑色）と Unique（青色）が色別で表されます。

ここでは青色はなく、緑色だけなので、4 つのブロックに対し全て Joint で構成されたモデルが作成されたことがわかります。



また、各ブロックのバー内に見える段数は、抽出された成分の数を表します。



例えば、1 番下の段は第 1 成分を表し、カーソルを対象となるブロックの成分上に持っていくと、そのブロックにおける R2 がポップアップされます。例えば Country C のブロックは 9 つの Joint 成分の合計が 0.703 となり、そのうち 0.152 が第 1 成分として寄与していることがわかります。

モデルの計算結果の解釈

Project ウィンドウのモデル行をダブルクリックし、Model ウィンドウを開きます。

Number	Model	Type	A	N	R2X(cum)	R2Y(cum)	Q2(cum)	Date	Title	Hierarchical
1	M1	MOCA	Autofitted	16				2019/12/11		

Model ウィンドウでは、モデルの詳細を見ることができます。

Model type: MOCA, observations (N): 16, variables (K): 556 (Judges=102, Country A=113, Country B=113, Country C=228)

Title:

Component	R2X	R2X(cum)	R2X	R2X(cum)	R2X	R2X(cum)	R2X	R2X(cum)
Model		0.778		0.681		0.662		0.703
Blocks	Judges		Country A		Country B		Country C	
Joint components		0.778		0.681		0.662	①	0.703
1	0.245	0.245	0.144	0.144	0.18	0.18	0.152	0.152
2	0.0596	0.305	0.211	0.355	0.129	0.309	0.134	0.286
3	0.134	0.439	0.0928	0.448	0.115	0.424	0.101	0.387
4	②	0.0686	0.507	0.0553	0.503	0.0558	0.48	0.069
5	0.0755	0.583	0.066	0.569	0.0571	0.537	0.055	0.511
6	0.145	0.728	--	--	0.0704	0.608	0.0941	0.605
7	--	--	0.0664	0.636	0.0544	0.662	0.0521	0.657
8	0.0498	0.778	0.0455	0.681	--	--	0.0459	0.703
Unique components		0		0		0		0

Global Joint (全共通変動)

Local Joint (一部共通変動)

Joint（上）と Unique（下）において、それぞれの成分で得られた、ブロック毎の R2 と累積 R2(cum)を表示します。

例えば、先程と同様に Country C のブロックを見てみると、全 8 成分で構成された Joint の 0.703 の内、第 1 成分では 0.152 の R2 が得られていることがわかります（図中①）。

また、第 6 成分では、Country A が成分として抽出されていないから、この成分は Judges、Country B、C での一部共通変動（Local Joint）であることがわかります（図中②）。

つまり、Joint の内、最初の 5 成分までは全ブロックでの共通変動（Global Joint）、6～8 成分は一部共通変動（Local Joint）として抽出されました。

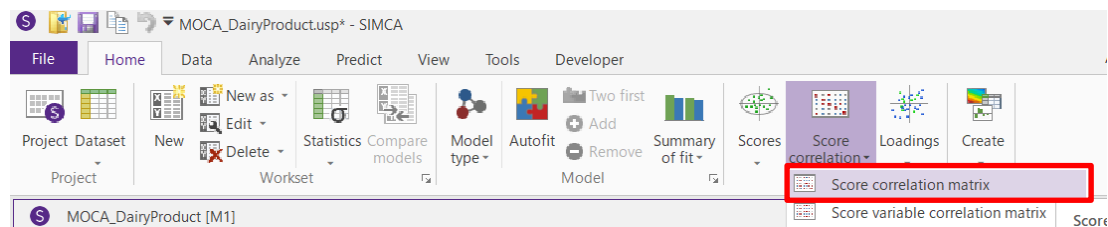
このように Model ウィンドウでは、各ブロック、各成分の情報を一覧で見ることが可能です。

スコア相関行列

スコア相関行列（Score correlation matrix）は、モデルの品質をチェックするためのツールです。

MOCA にのみ搭載されています。

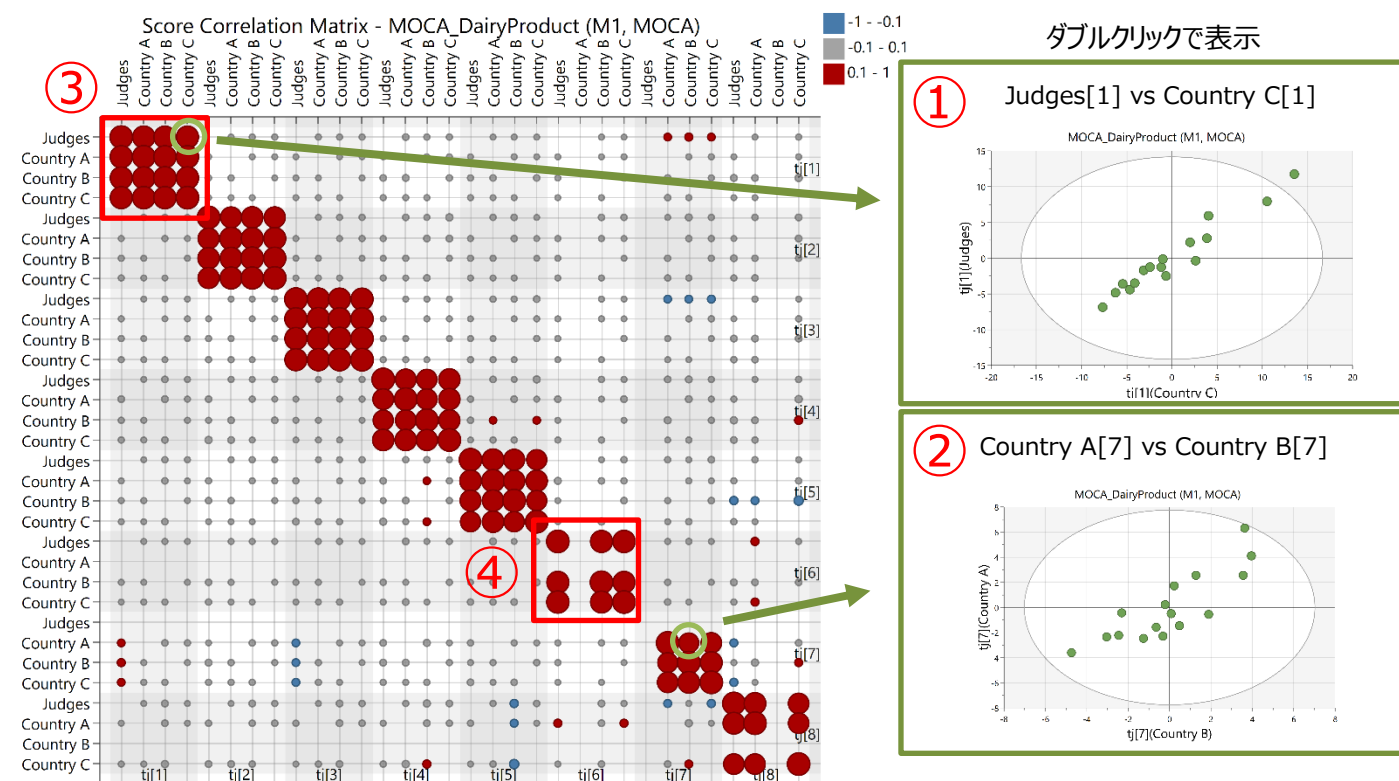
Home タブ > Score correlation > Score correlation matrix で表示します。



表では Joint と Local の各成分における、ブロック間の相関関係を表します。

赤色は正の相関、青色は負の相関を表し、丸の大きさは、相関の強さを表します。（図①、②）

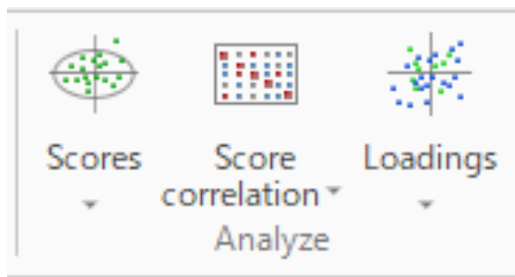
左上の赤枠③は第 1 成分における 4 ブロックの相関関係を表し、④は第 6 成分における 4 ブロックの相関関係を表します。③では正方形の中の全てで相関があるので、全ブロックで共通の変動（Global Joint）を持っていることが確認できます。④では Country A は Joint の成分としては検出されず、残りの 3 ブロックでのみ相関を持つ、一部共通変動（Local Joint）を持っていることがわかります。このことは Model ウィンドウでも確認できました。



斜めの対角線上以外で強い相関が確認できる場合、そのブロックに大きな外れ値が含まれている可能性があり、さらなる調査が必要です。ここではそのような相関は見られないので、良いモデルが得られていると判断できます。

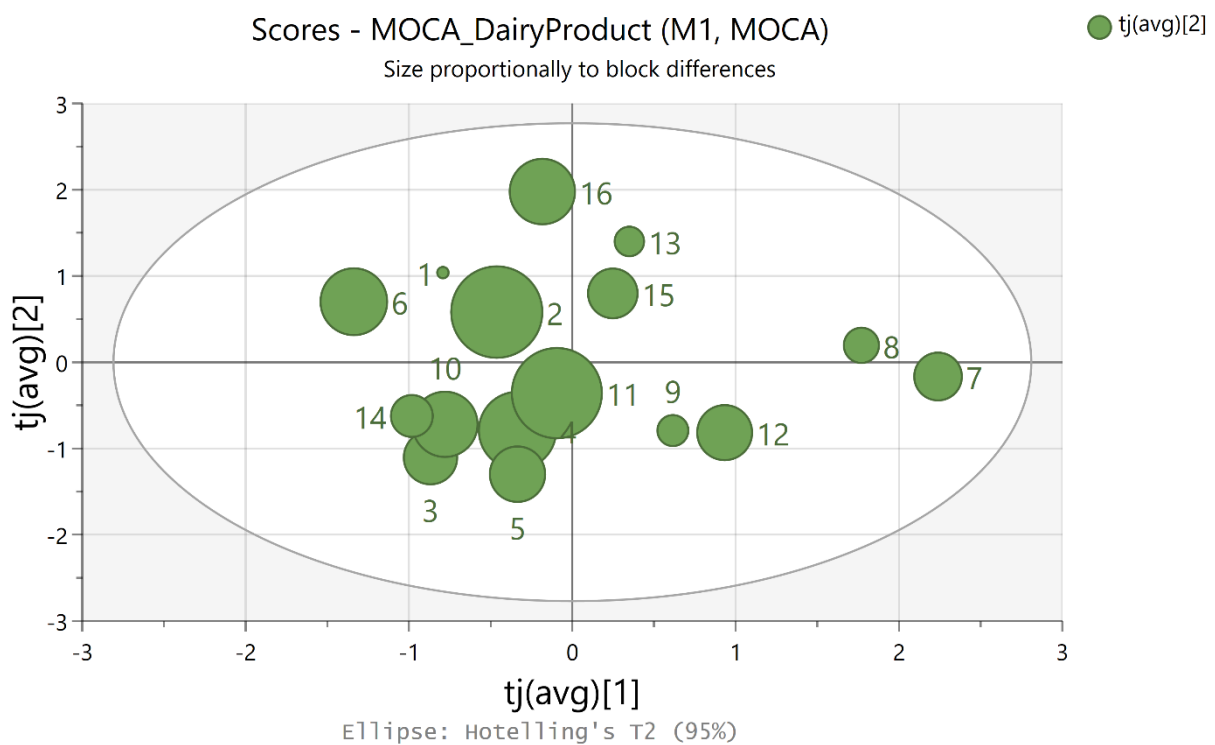
モデルの表示

MOCA では、以下のリボンが表示されます。



スコアプロット

Scores をクリックし、スコアプロットを表示します。

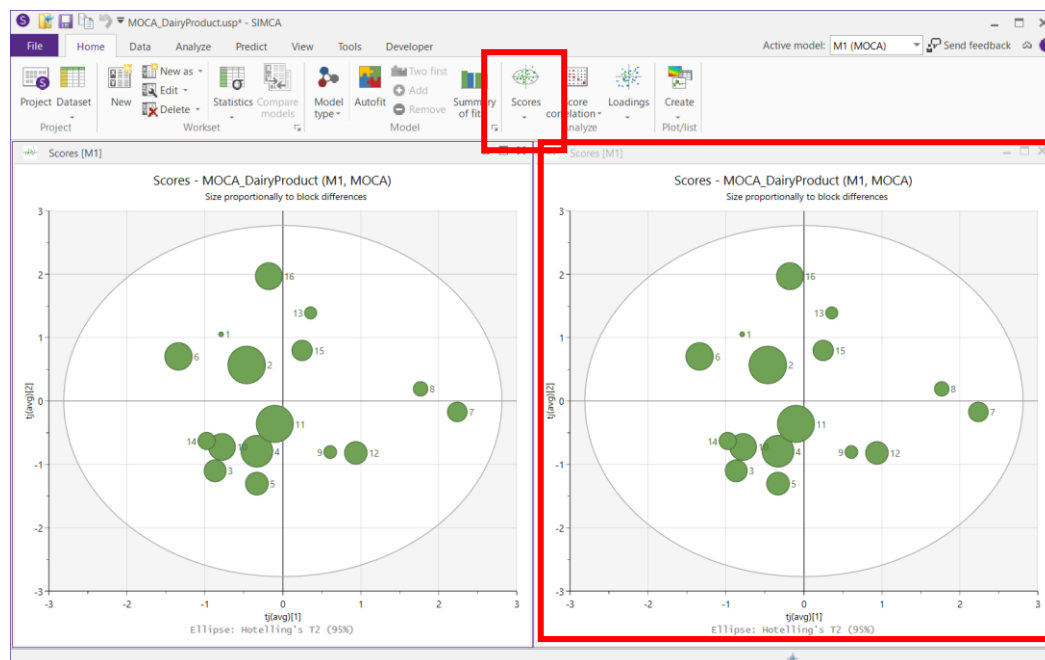


MOCA では、デフォルトで全ブロックの Joint スコアの平均 (tj(avg)) が表示されます。

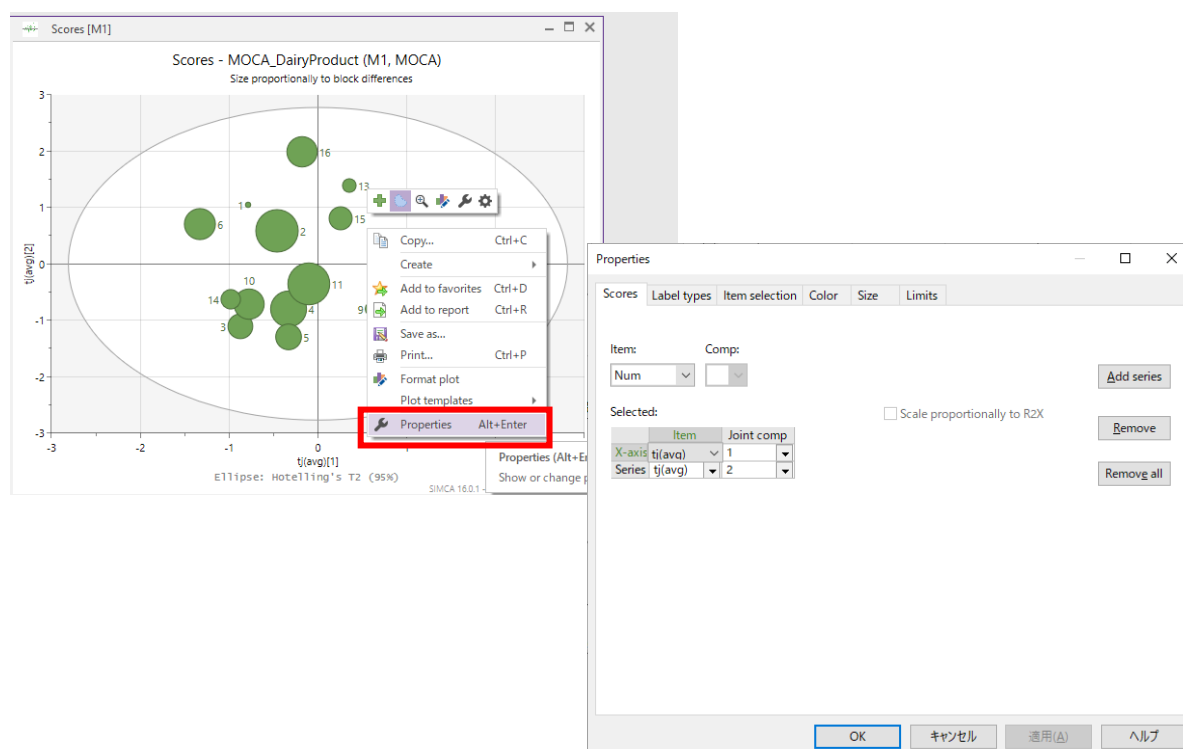
丸の大きさは、ブロック間での、得られたサンプルのスコアの違いを表します。小さい丸は、各ブロックで同じ位置にサンプルのスコアが得られ、大きい丸は、サンプルのスコアの位置が異なるブロックが得られていることを示します。

各ブロックのスコアプロットを表示し、ブロック間のスコアの違いを確認します。

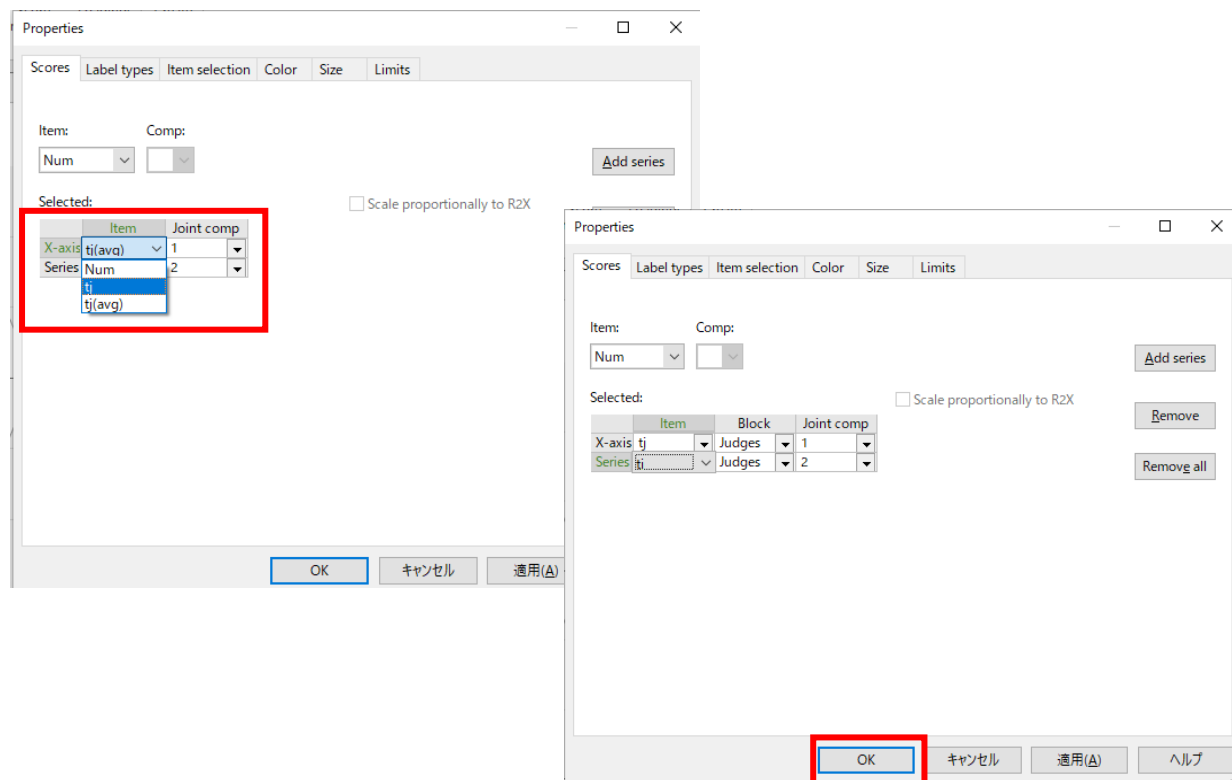
もう一度 Home タブ > Scores をクリックし、スコアプロットを表示します。



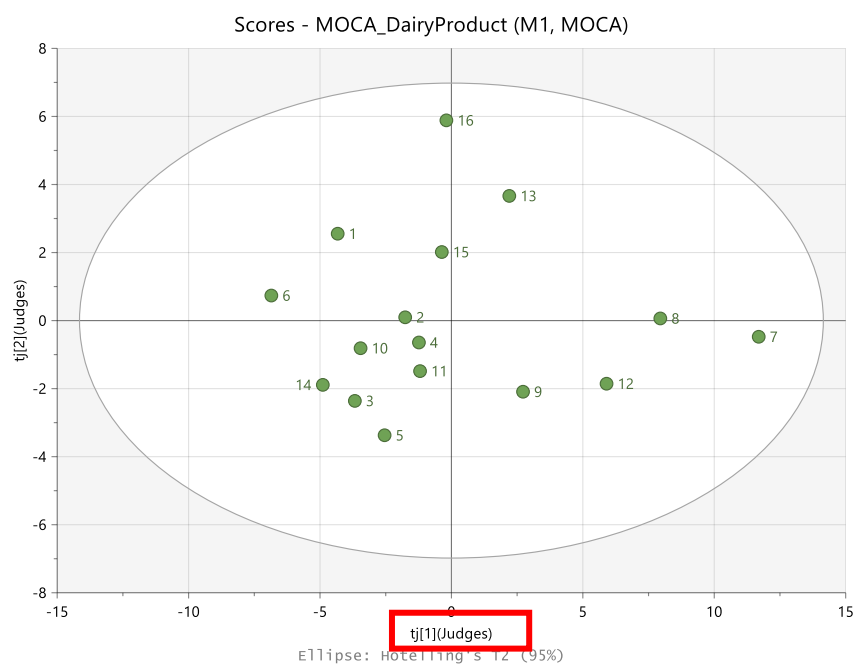
新たに表示したスコアプロット上で右クリック → Properties を選択し、Properties 画面を開きます。



Item 列の X-axis, Series を tj(avg) → tj に変更し、Block の X-axis, Series を Judges に変更します。
変更後、OK をクリックします。（※Block 列 は tj に変更した際に、自動で追加されます。）

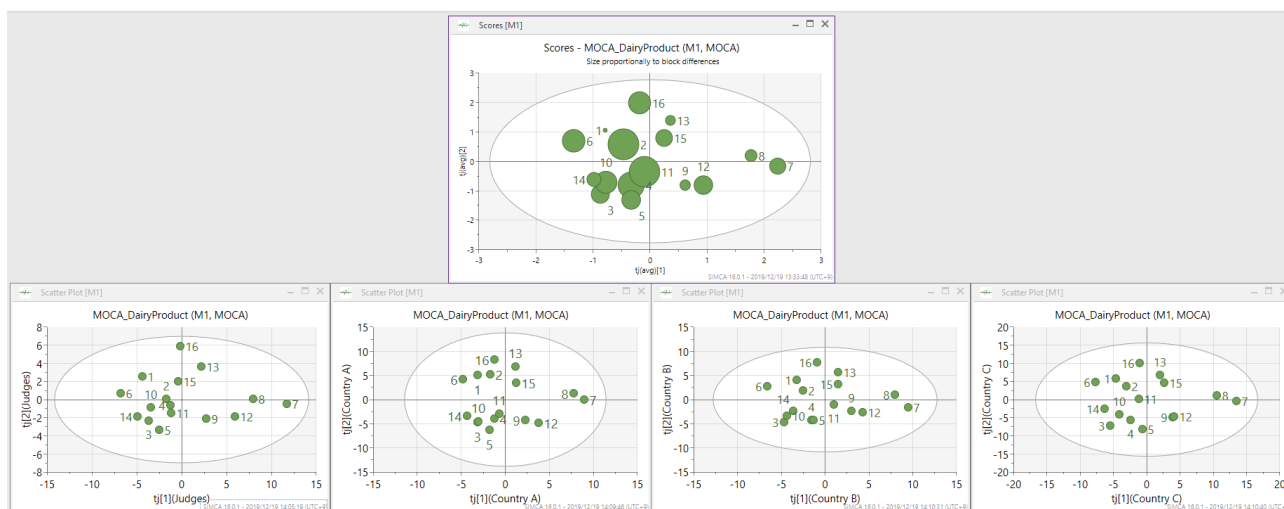


Judges の単独のスコアプロット(tj)が得られました。

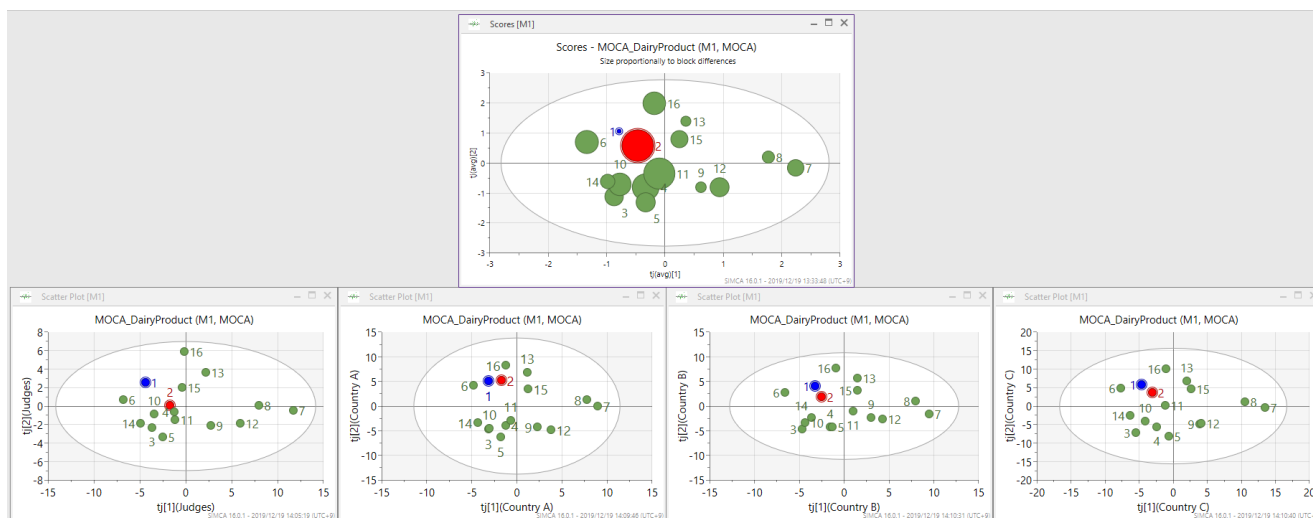


新たにスコアプロットを表示し、先程と同様の手順で、Block のみを変更し、残り 3 ブロックについて単独のスコアプロットを表示します。

下図のように全ブロックのスコアプロットと、各ブロックのスコアプロットが得られます。



全ブロックスコアプロット上の 1、2 を順番にクリックします。

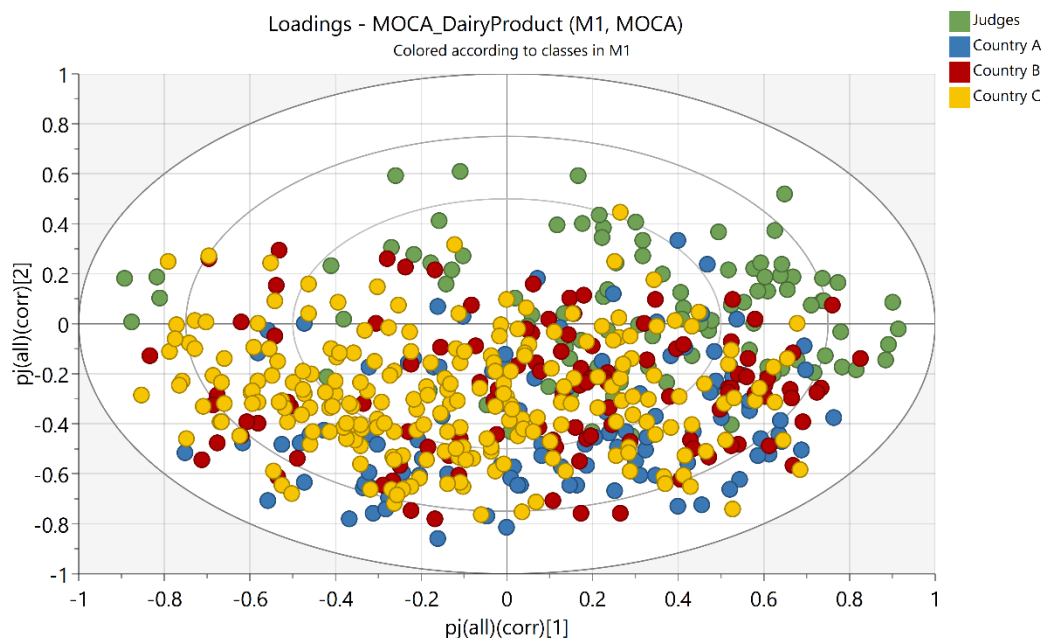


大きさの小さい 1 は、どのブロックでも同様のスコアが得られ、
大きさの大きい 2 は、ブロック間ではばらつきがあることがわかります。

ローディングプロット

Home > Loadings をクリックし、ローディングプロットを表示します。

スコアプロットと同様に、全ブロックの Joint ローディングの平均 (pj(all)) が表示されます。



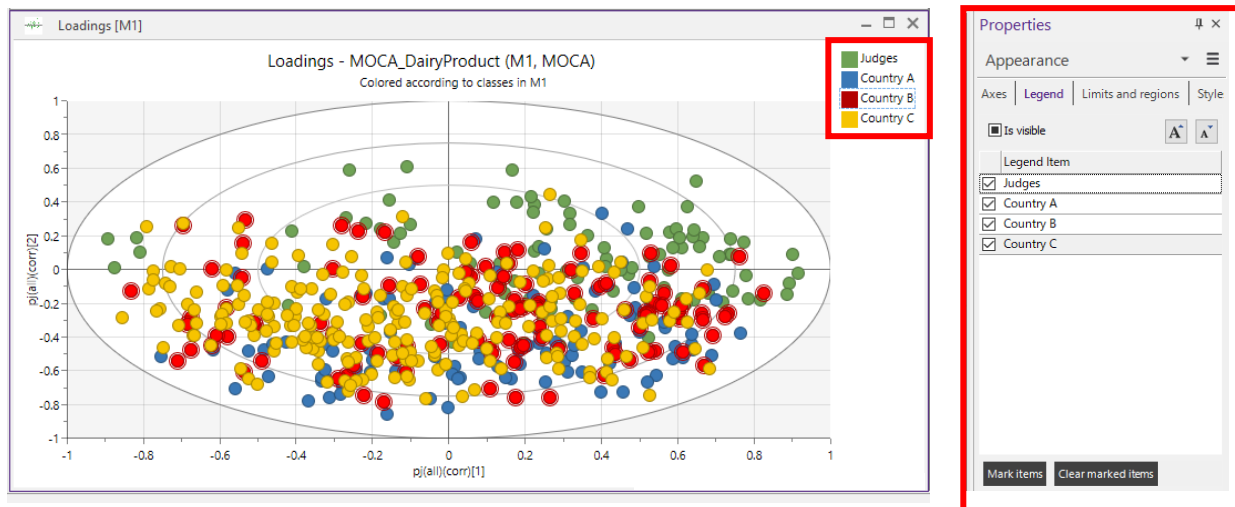
スコアプロットと対比させるために、ローディングプロットも同様にブロックごとに表示させます。

以下では凡例 (Legend) を利用して表示させます。(スコアプロットと同じ方法でも問題ありません)

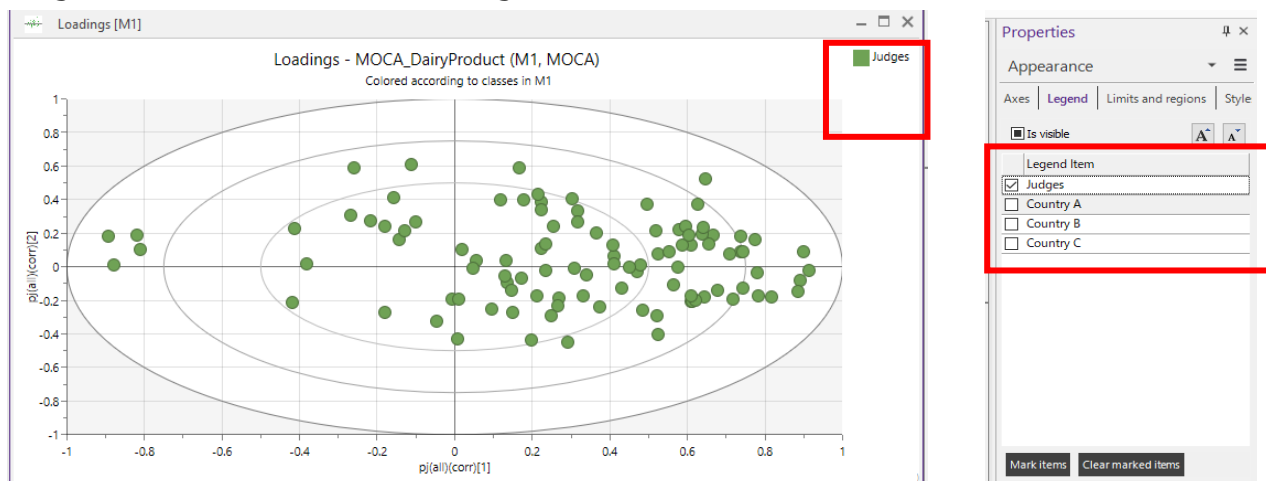
Home タブ > Loadings を 4 回クリックし、新たに 4 つローディングプロットを表示します。



変更したいローディングプロットの画面右上の凡例をクリックすると、Explorer 画面が Appearance に変更されます。
(または、Properties タブ > Appearance > Legend)



Judges 以外のチェックボックスを外すと、Judges のみが表示されます。



同様の手順で、Country A,B,C のみにチェックをいれ、表示します。

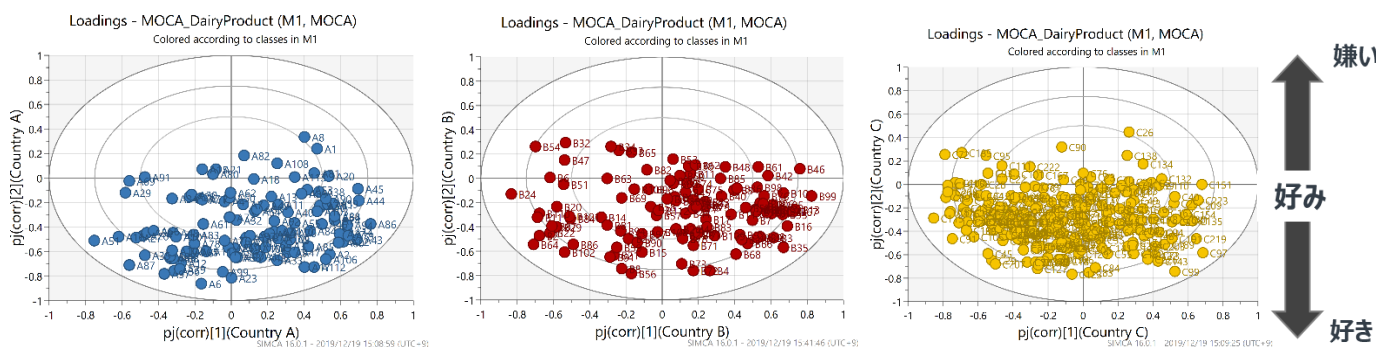


ローディングプロットから、Judges は右半分、A,B,C の 3 国は下半分に分布が偏っているのがわかります。

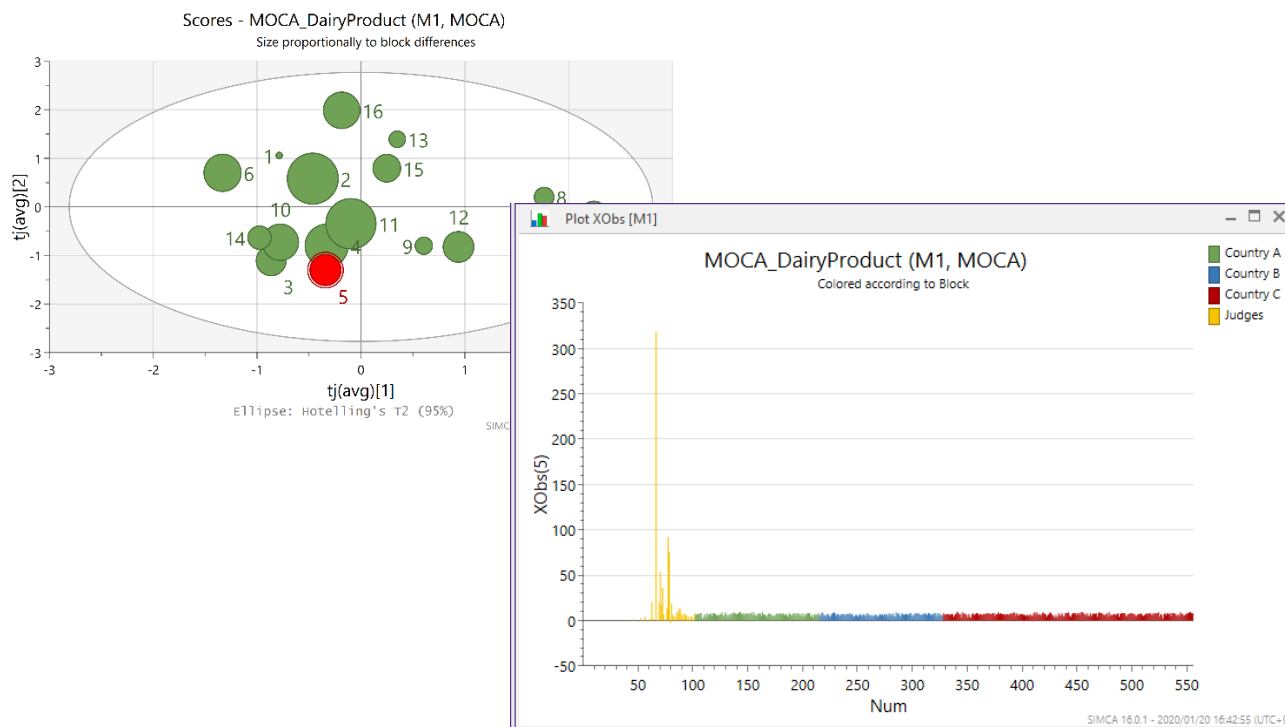
3. 結果を解釈する

MOCA モデルの結果を解釈し、消費者に好まれる製品の特徴を探していきます。

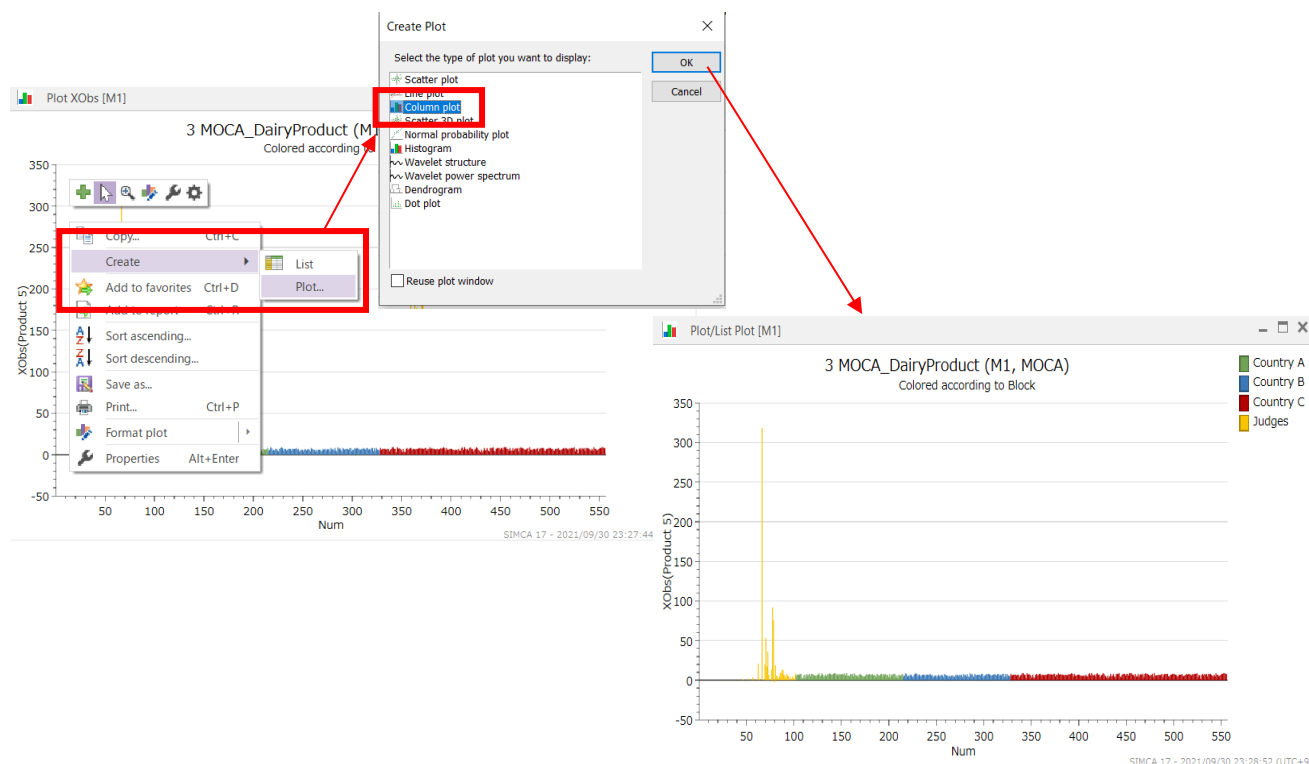
A,B,C の3国は下半分に分布が偏っており、縦軸が好みを表す軸であることがわかります。
このことから、スコアプロットの下の方に位置する製品は好まれる傾向にあることがわかります。



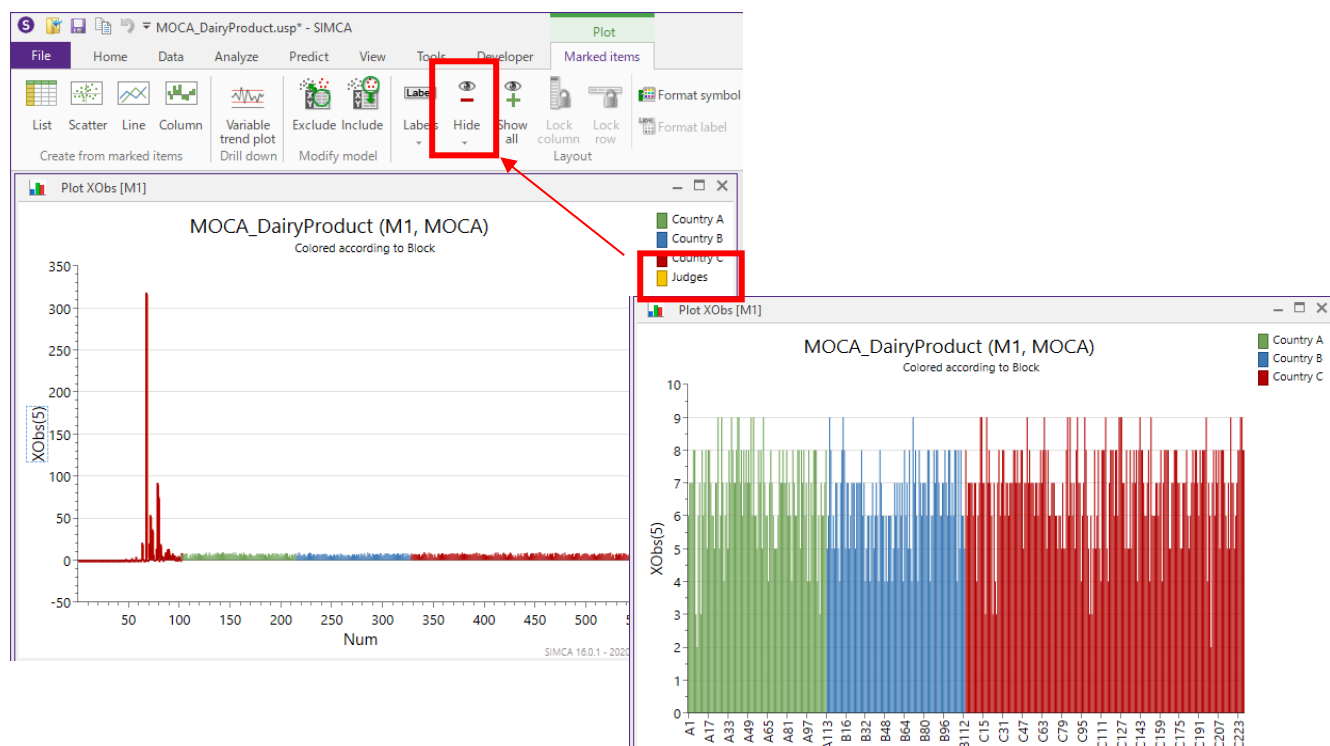
検証していきます。スコアプロットの5をダブルクリックし、Plot XObsを表示します。



表示したプロット上で右クリック > Create > Plot を選択。Column plot をクリックし OK すると、同様のプロットが表示されます。



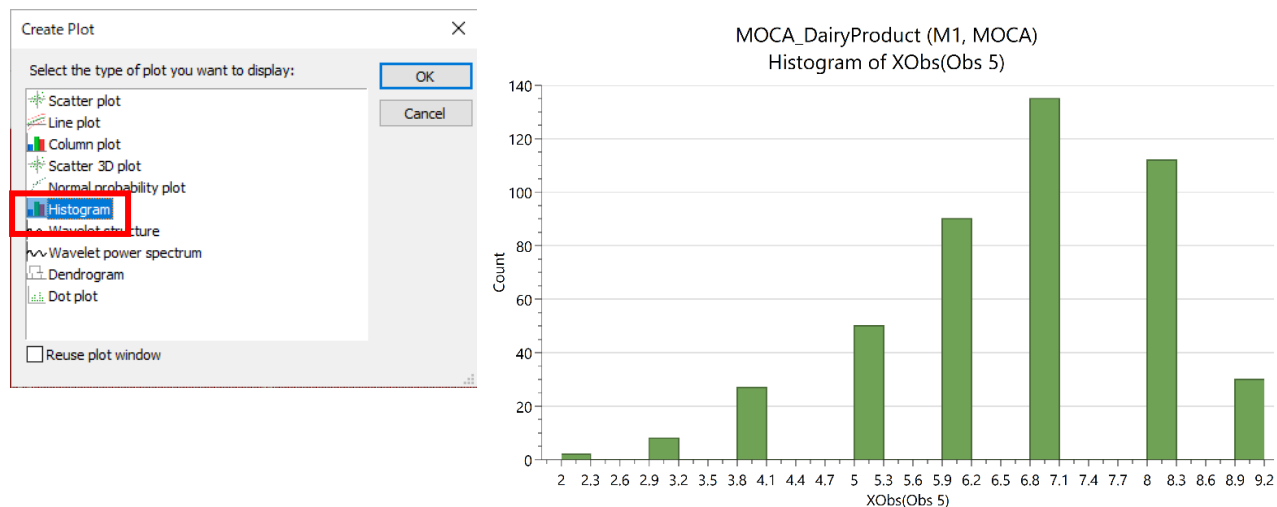
凡例 (Legend) の Judges をクリックし、Marked item タブ > Hide をクリックします。
Judges はグラフから外され、Country A,B,C のみのグラフが表示されます。



製品 5 における 3 国の Xobs が表示されます。消費者による 9 点評価のグラフなので、最大値は 9 となります。

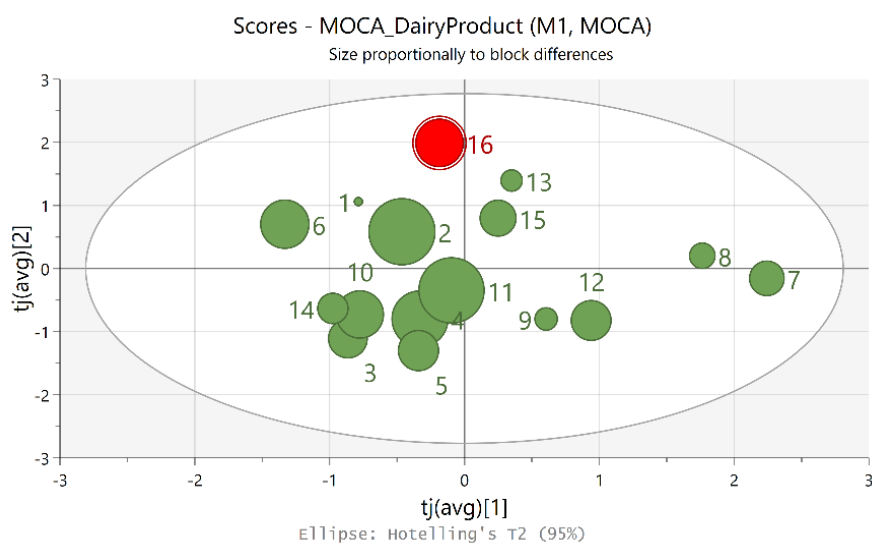
このグラフ上で右クリック > Create > Plot にて、Histogram を選択し OK します。

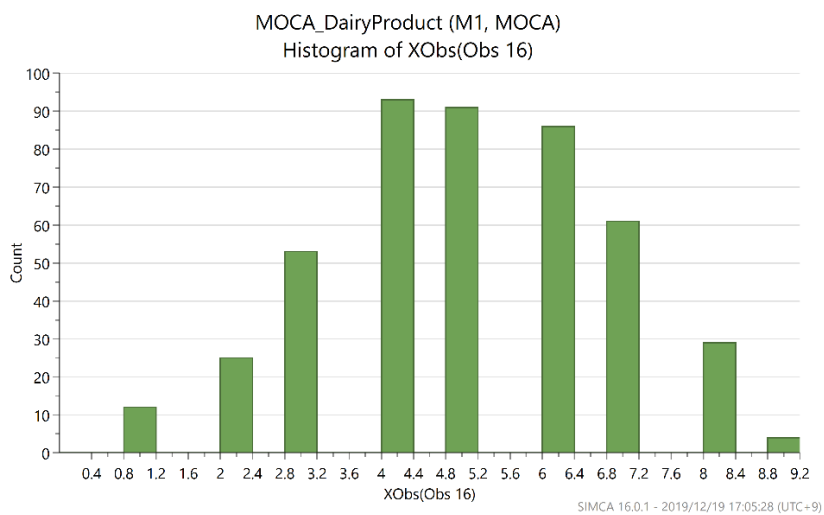
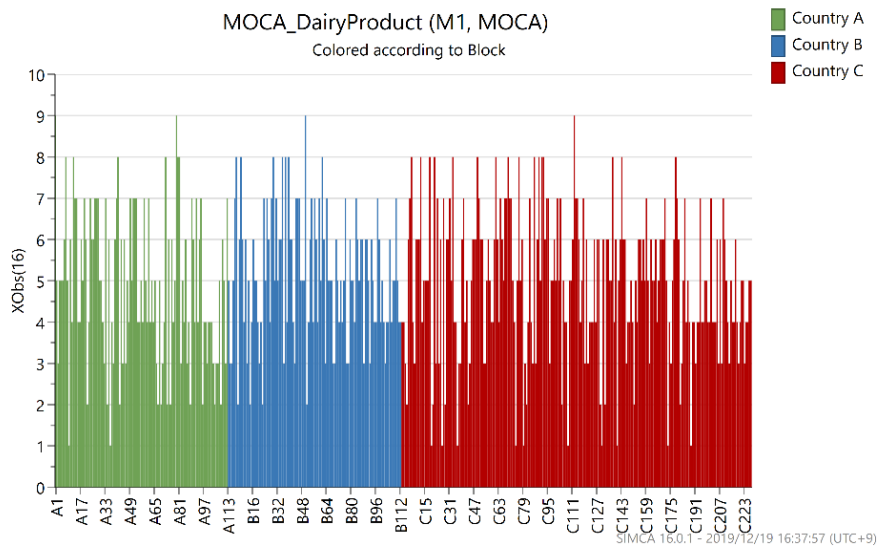
製品 5 における 3 国の消費者の得点分布がわかります。



製品 5 では、7 点あたりを中心に、高めの点数分布であることがわかります。

同様の手順で製品 16 についても、Xobs、および Histogram を表示します。

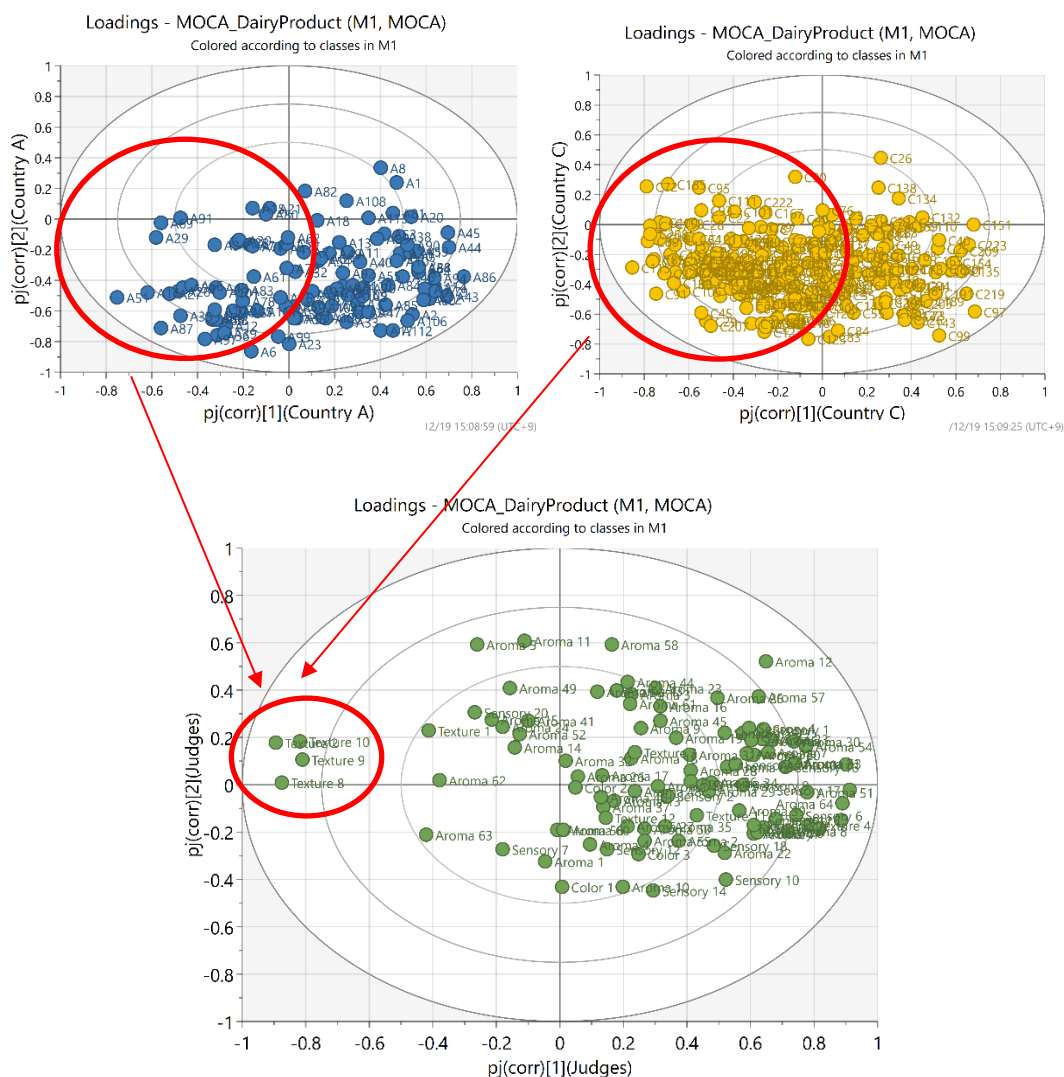




製品 16 では 4 点付近の点数分布であることがわかります。

先程の製品 5 の結果との比較から、縦軸下に位置する 5 は、消費者に好まれ、16 は好まれない傾向にあることがわかります。

3 国ともに縦軸が製品の好みを表すことがわかりましたが、A 国、C 国では左右の分布の差があります。
この国間の違いは、官能評価のローディングプロットで説明ができます。



官能評価で左側に分布している Texture2,8,9,10 は、左側の分布が少ない A 国では好まれず、分布の多い C 国では好まれる傾向であることがわかります。

同様に右側の官能評価の項目に関しては、A 国に好まれ、C 国には好まれる傾向であることがわかります。

このことから、横軸は国間の違いを表す軸である、とも言い変えることができます。

4. 最後に

本チュートリアルでお示した解析例は、MOCA 解析のごく一部です。

以下の参考文献などをご参考ください。

・熱帯熱マラリアに感染した子供の血漿サンプルから得られたリポミクス、メタボロミクス、オキシピリンデータのマルチブロックデータ解析

Izabella S, et al., Faraday Discuss., 2019, 218.

“Joint and unique multiblock analysis of biological data – multiomics malaria study”

・喘息コホート患者のトランスクリプトーム、メタボロミクス、脂肪酸、スフィンゴ脂質、オキシリピン、臨床データによるマルチブロックデータ解析

Stacey R, et al., Faraday Discussions. 2018, 90, 13400–13408

“OnPLS-Based Multi-Block Data Integration: A Multivariate Approach to Interrogating Biological Interactions in Asthma”

・5 つの NIR 機器で測定したマッシュルームサンプルのマルチブロックデータ解析による、機器の類似点と相違点の調査

Skotare T, et al., Anal.Chem, 2019,91,3516-3624.

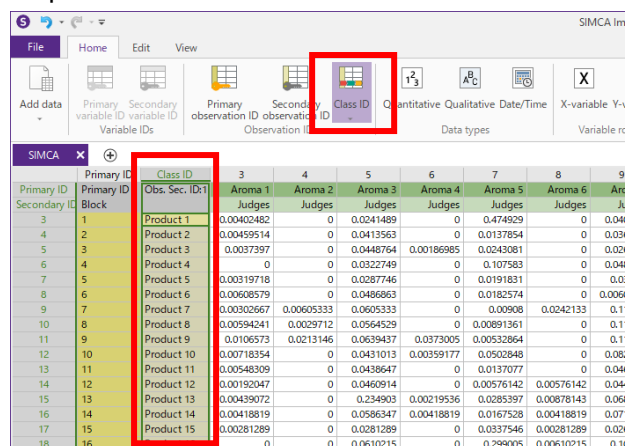
“Joint and Unique Multiblock Analysis for Integration and Calibration Transfer of NIR Instruments”

◆既知の問題

Class ID を設定すると、MOCA モデルの作成に必要な「Block」の設定ができなくなります。

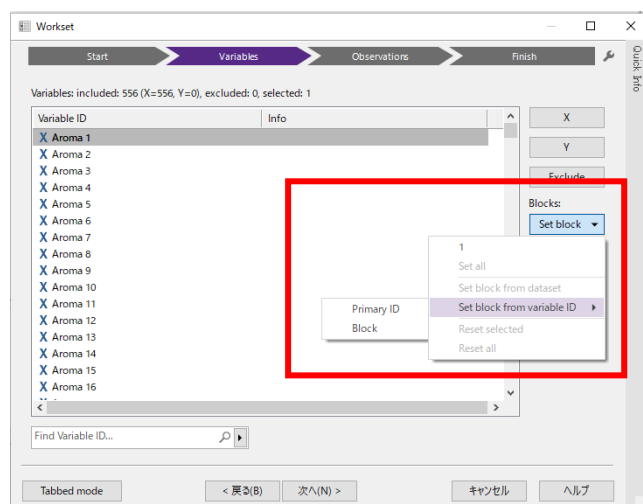
■症状

Import 画面で、Class ID を設定。

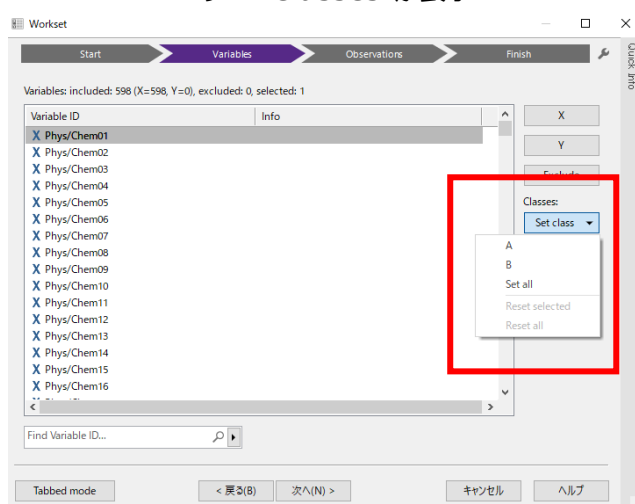


Workset > Variables の設定画面で “Block” の選択が消え、“Classes” が表示されています。

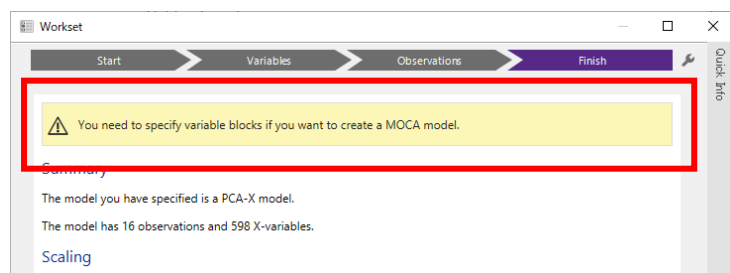
通常：Blocks が表示



エラー：Classes が表示



Variables は Class 設定の必要はないため、特に設定をせず Workset を進めると、Block の設定がされていないため MOCA モデルが作成できない、というエラーが表示されます。



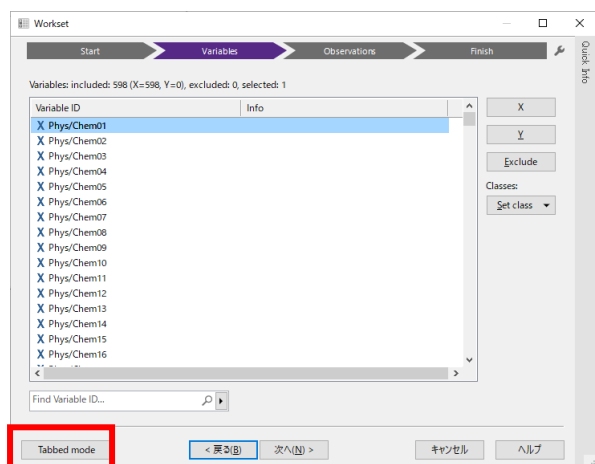
■ 対処法

① Class ID を使用しない

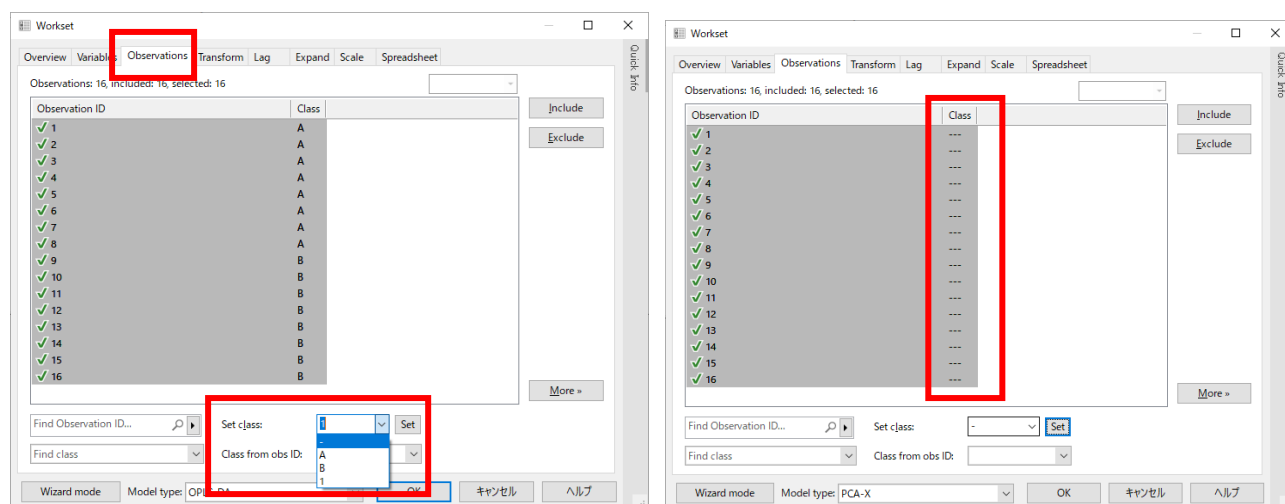
最初のデータインポート画面にて、Class ID を使用せず、Secondary ID を設定していれば、このような症状は起こりません。そのため、MOCA では Class ID を設定されないことを推奨します。

② (Class ID を設定した場合) Tabbed mode 画面にて、Class 設定を解除する

Workset の左下 Tabbed mode をクリックし Tabbed モードの画面に変更します。



Observations タブにて、Ctrl + A で全ての Observation を選択、Set classes を "-" に変更し、Set をクリックします。クラス情報が無効にされたことを確認します。



これにより、Variables タブの Block 設定ボタンが復活します。

このまま Tabbed モードのままでも、Wizard モードに戻しても、設定を進めることが可能です。

