

# CONTENTS

---

<b>サンプルの Experiment – PCA Demo</b> .....	2
クイックノート .....	2
デモの操作手順.....	2
PCA コマンドのヘルプ .....	4

# サンプルの Experiment – PCA Demo

## クイックノート

メニュー **File** → **Example Experiments** → **Analysis** → **PCA Demo**

この Experiment では、組み込みの PCA（主成分分析）コマンドと外部プロシージャの使い方を説明します。

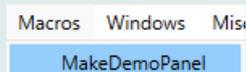
この Experiment には必要なプロシージャがすべて含まれているので、ご自身のデータを分析する際の出発点としても使うことができます。

PCA のコマンドおよびプロシージャは、E.R. Malinowski 著『Factor Analysis in Chemistry』（John Wiley and Sons、1991 年）に従っています。

オリジナルの Igor プロシージャは、1999 年に David W. Niles によって書かれました。

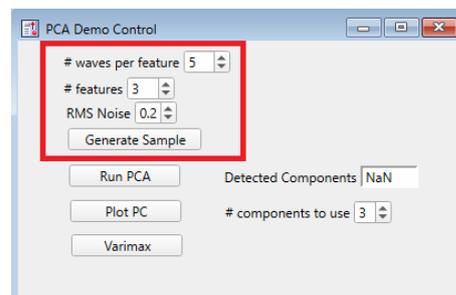
## デモの操作手順

**1. デモを開始するには、Macros メニューから MakeDemoPanel を選択します。**



**2. PCA Demo Control パネルの上部には、分析用のサンプルデータを作成するためのコントロールがあります。**

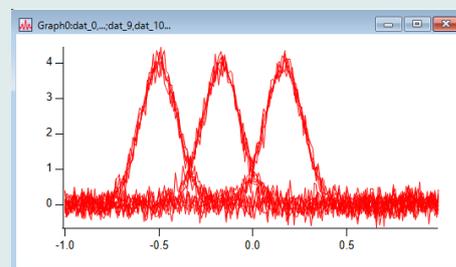
サンプルデータは、ノイズを加えたガウスプロファイルの集合で構成されており、「feature」は異なる線の中心位置に対応しています。すべてのガウス曲線は、同じ線幅とノイズレベルで生成されます。生成されるウェーブの総数は、feature の数と feature ごとのウェーブの数の積です。



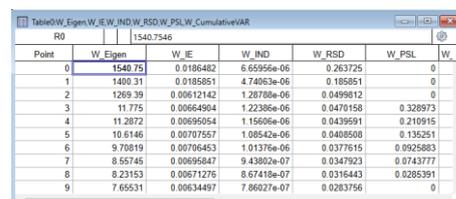
**3. 次のように設定して、Generate Sample ボタンをクリックします：**

# waves per feature: 5  
# of features: 3  
RMS Noise: 0.2

右図のようなグラフが作成されます。



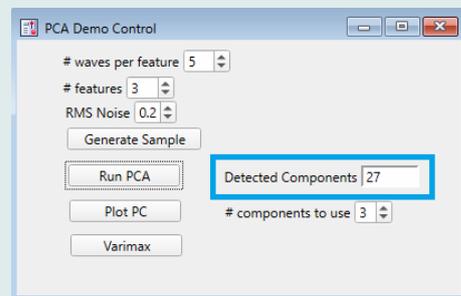
**4. サンプルデータを生成した後、Run PCA ボタンをクリックすると、ウェーブの集合に対して主成分分析のステップが実行されます（実行には少し時間がかかるかもしれません）。**



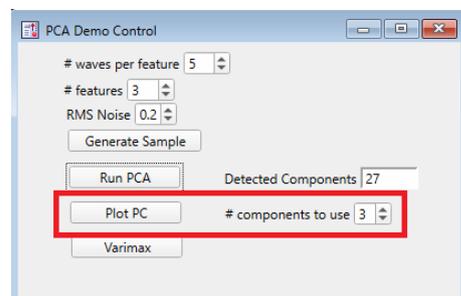
Point	W_Eigen	W_IE	W_IND	W_RSD	W_PSL	W
0	1540.75	0.0186482	6.65956e-06	0.263725	0	
1	1400.31	0.0185851	4.74063e-06	0.185851	0	
2	1269.39	0.00612142	1.28788e-06	0.0499812	0	
3	11.775	0.00664904	1.22386e-06	0.0470158	0.328973	
4	11.2072	0.00695054	1.15006e-06	0.0439591	0.210915	
5	10.6146	0.00707557	1.08542e-06	0.0408508	0.135251	
6	9.70819	0.00706453	1.01376e-06	0.0377615	0.0925883	
7	8.55745	0.00695847	9.43802e-07	0.0347923	0.0743777	
8	8.23153	0.00671276	8.67418e-07	0.0316443	0.0285391	
9	7.65531	0.00634497	7.86027e-07	0.0283756	0	

5. 重要な成分の数はパネルに表示され、固有値、対応する埋め込み誤差 (W\_IE)、指示関数値 (W\_IND)、残差標準偏差 (W\_RSD)、有意水準 (W\_PSL) の数値結果をテーブルに表示します。

注: IE、IND、RSD、%SL の定義は Malinowski によるものです。

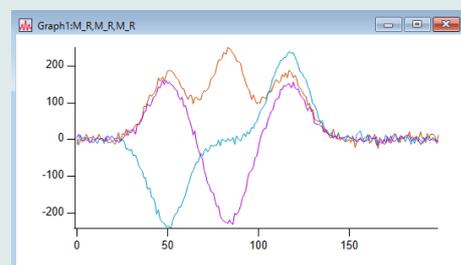


6. PCA を実行した後、パネルにその値を入力して表示したい成分の数を指定し、Plot PC ボタンをクリックすると、対応する成分が表示されます。

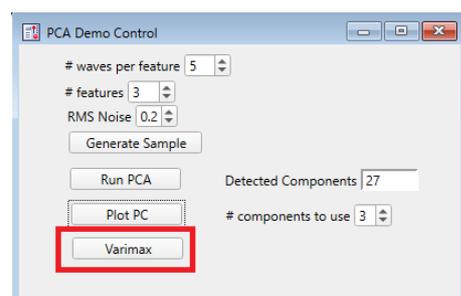


7. 分析の最終ステップは、Varimax 回転を実行することです。

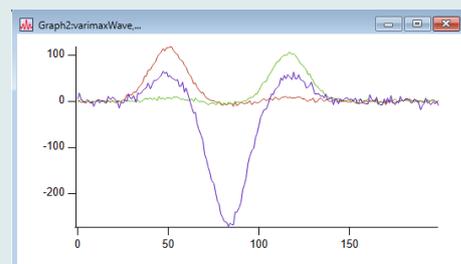
Varimax の計算は、Henry Kaiser の論文「Computer Program for Varimax Rotation in factor analysis」に基づいています: Educational and Psychological Measurement, Vol XIX, No. 3. 1959 PP. 413-420.



8. Varimax ボタンをクリックします。



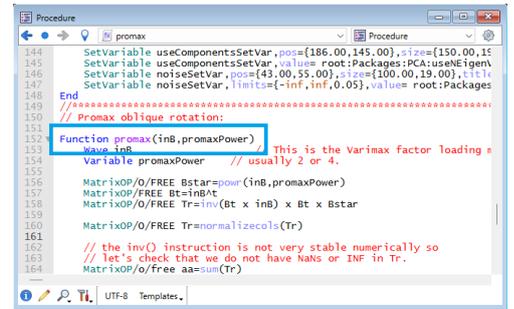
9. Varimax 回転は、前のステップで選択されたコンポーネントのみに影響します。



10. Promax 回転を試してみたい場合は、メインプロシージャウィンドウで promax() 関数を見つけることができます。

この関数は、Varimax プロシージャの出力と promax パワー変数を入力として受け取ります。

回転は、A.E. Hendrickson と P.O. White による論文「Promax: A quick method for the rotation to oblique simple structure」(「プロマックス: 斜交単純構造への迅速な回転法」) を基にしています。British Journal of Statistical Psychology, 17, 65-70ページ(1964年)。



```
144 SetVariable useComponentsSetVar, pos={186.00, 145.00}, size={150.00, 150.00}
145 SetVariable useComponentsSetVar, value= root:Packages:PCA:useEigenV
146 SetVariable noiseSetVar, pos={43.00, 55.00}, size={100.00, 19.00}, titl
147 SetVariable noiseSetVar, limits={-inf, inf, 0.05}, value= root:Packag
148 End
149 // Promax oblique rotation:
150
151
152 Function promax(inB, promaxPower)
153   /* inB // This is the Varimax factor loading e
154   Variable promaxPower // usually 2 or 4.
155
156   MatrixOP/O/FREE Bstar=power(inB, promaxPower)
157   MatrixOP/FREE Bt=inB*Bt
158   MatrixOP/O/FREE Tr=inv(Bt x inB) x Bt x Bstar
159
160   MatrixOP/O/FREE Tr=normalizecols(Tr)
161
162   // the inv() instruction is not very stable numerically so
163   // let's check that we do not have NaNs or INF in Tr.
164   MatrixOP/o/free aa=sum(Tr)
```

## PCA コマンドのヘルプ

```
PCA [/ALL/COV/CVAR/IE/IND/LEIV/NF/O /Q /RSD[=rsdMode] /RMS/SCMT  
/SCR/SDM/SEVC/SL/SQEV/SRMT/U/VAR/WSTR=waveListStr /Z] [wave0, wave1, wave2 ... wave99]
```

PCA コマンドは、主成分分析を実行します。

入力データは、1D ウェーブのリスト、単一の 2D ウェーブ、または 1D ウェーブのリストを含む文字列の形式で入力できます。

このコマンドでは、指定されたフラグに応じて複数の出力ウェーブが生成されます。

### フラグ

- /ALL** よく使われるフラグの組み合わせのショートカット: /CVAR、/SL、/NF、/IND、/IE、/RMS。
- /COV** 入力ウェーブ (複数可) 共分散行列を計算し、残りの分析の入力として使用します。共分散行列は、まず各入力 1D ウェーブを連続する列にコピーした行列を作り、その行列をその転置行列で乗算することで計算されます。
- /CVAR** 累積パーセント変動を計算します。これは、 $100 * (\text{最初の } m \text{ 個の固有値の合計をすべての固有値の合計で割った値})$  として定義されます。結果は、現在のデータフォルダー内のウェーブ W\_CumulativeVAR に保存されます。下記の /VAR も参照してください。
- /IE** 埋め込み誤差を計算します。現在のデータフォルダー内のウェーブ W\_IE で誤差を返します。ウェーブは、SetScale/P x 1,1,"", W\_IE を使ってスケールされます。埋め込まれた誤差は、要因の数、行と列の数、有意因子に含まれない固有ベクトルの合計の関数です。IE の動作は、有意因子の数を決定します。
- /IND** 因子インジケータ関数を計算します。/IND を指定すると、残差標準偏差も計算されることに注意してください。現在のデータフォルダー内のウェーブ W\_IND に結果を返します。ウェーブは、SetScale/P x 1,1,"", W\_IND を使ってスケールされます。
- /LEIV** SVD (特異値分解) 計算にあまりメモリを必要としないように、固有値を制限します。上限は、入力データの行または列の最小数に設定されます。

/NF 有意因子の数を求め、変数 V\_npnts に格納します。  
有意因子を計算するには、/IND を使う必要があります。

/O 入力ウェーブを上書きします。

/Q コマンドウィンドウの履歴領域に結果を表示しません。

/RSD[=rsdMode] 残留標準偏差 (RSD) を計算し、現在のデータフォルダー内のウェーブ W\_RSD に RSD を返します。  
W\_RSD の最初の要素は NaN であり、残りのすべてのウェーブ要素は有意因子の数に対応しています。  
原点に関する共分散の場合は、rsdMode=0 を使います。  
原点に関する相関の場合は、rsdMode=1 を使います。

/RMS RMS の誤差を計算します。  
現在のデータフォルダー内のウェーブ W\_RMS に結果を返します。  
ウェーブは、SetScale/P x 1,1,"", W\_RMS を使ってスケールされます。

/SCMT 現在のデータフォルダー内のウェーブ M\_C に特異値分解 (SVD) 後の C 行列を保存します。

/SCR 個々のウェーブ入力を標準スコアに変換します。  
入力が単一の 2D ウェーブの場合には、動作しません。  
ウェーブの 1 つ以上のエントリが NaN または INF の場合に標準スコアに変換するのは誤りです。  
この機能を使う場合は、RSD 計算の適切な形式を使用してください。

/SDM 計算終了時にデータマトリックスのコピーを保存します。  
入力が個々のウェーブで構成されている場合や、計算された標準スコアを保存したい場合に便利です。  
入力が 2D 行列の場合、入力行列のコピーがウェーブ M\_D に入ります。

/SEVC 固有値ベクトルをウェーブ W\_Eigen に保存します。  
これは、SVD によって生成された生の固有値であり、現在のデータフォルダーに保存されます。  
通常、SVD が共分散または相関行列ではなく、生のデータ行列に適用された場合、PCA 固有値を求めるには、各ウェーブの要素を 2 乗する必要があります。  
このウェーブにはデフォルトのウェーブスケールが適用されていることに注意してください。

/SL 有意水準のパーセンテージを計算し、現在のデータフォルダー内のウェーブ W\_PSL に格納します。

/SQEV SVD 固有値の平方和を求めません。  
/COV を指定した場合は、このフラグを使う必要はありません。  
入力がすでに共分散行列である場合のみ使ってください。  
その場合、SVD の結果は固有値であり、その平方根ではありません。

/SRMT SVD 後の R 行列を保存します。

/U 入力が 2D ウェーブの場合のみ、入力ウェーブを変更せずに残します。  
適切なフラグが使われている場合でも、共分散の計算は行われないうことに注意してください。

/VAR 各固有値に関連する分散を計算します。  
分散は、固有値の総和に対する固有値の比率として定義されます。

結果は、現在のデータフォルダー内のウェーブ W\_VAR に保存されます。  
上記の /CVAR も参照してください。

/WSTR=waveListStr  
すべての入力ウェーブの名前の文字列リストです。

/Z エラーを無視します。

## 詳細

入力は、/WSTR=waveListStr、最大 100 個の 1D ウェーブのリスト、または最後のフラグに続く単一の 2D ウェーブのいずれかです。

waveListStr は、データマトリックスに使う 1 次元ウェーブのセミコロン区切りのリストを含む文字列です。

waveListStr には、ウェーブへの適切なパスを任意で含めることができます。

リベラルな名前は引用符で囲むことも、囲まないこともできます。

すべてのウェーブは同じ数値型（単精度または倍精度）であり、すべてのウェーブのポイント数は同じであると想定されています。

入力に関わらず、この操作では結果の行列の行数が列数以上であることを想定しています。

コマンドは、入力されたウェーブからデータマトリックスを作成することから始まります。

1D ウェーブのリストを提供すると、それらがデータマトリックスの列になります。

共分散行列 (/COV) をデータ行列として使うこともできます。

また、データ行列の各列を標準スコアに変換するために正規化することもできます。

これは、各列の平均と標準偏差を計算し、その後新しい値を次のように設定することを意味します。

$$\text{newValue} = (\text{oldValue} - \text{colAverage}) / \text{colStdv}$$

入力データは、SubtractMean、NormalizeRows、NormalizeCols 関数を使って MatrixOp で前処理することができます。

データマトリックスを作成した後、その操作ではデータマトリックスの特異値分解 (SVD) を計算します。

SVD の結果は保存したり、さらに処理したりすることができます。

C 行列と R 行列は、/SCMT および /SRMT を使って保存します。

これらは、次を通して入力データ行列に関連しています：

$$D = R * C$$

コマンドの残りの部分では、Malinowski によって定義されたさまざまな統計量を計算することができます（参考文献を参照）。

フラグを使って、どのフラグが計算されるかを決定します。

このコマンドにより、いくつかの出力ウェーブが生成されます。

すべてのウェーブは現在のデータフォルダーに保存されます。

入力行列 D はウェーブ M\_D に保存でき、オプションの SVD 結果は、列行列 C を含むウェーブ M\_C、

行行列 R を含むウェーブ M\_R、およびデータ行列の固有値を含む W\_Eigen に保存されます。

これらは、入力行列が共分散行列であるか否かによって、固有値または固有値の 2 乗となることに注意してください (/SQEV フラグを参照)。

オプションの 1D 出力ウェーブ (W\_RSD、W\_RMS、W\_IE、W\_IND、W\_PSL) は、ウェーブのスケーリングとともに保存され、因子数の関数としてウェーブを表示しやすくします。

## 参考

Kaiser, H., Computer Program for Varimax Rotation in Factor Analysis, Educational and Psychological Measurement, XIX, 413-420, 1959.

Malinowski, E.R., Factor Analysis in Chemistry, 3rd ed., John Wiley, 2002.