

CONTENTS

Volume II User's Guide: Part 1 (Graphs #6)	2
ボックスプロットとバイオリンプロット	2
ボックスプロットとバイオリンプロットの文法	2
ボックスプロットとバイオリンプロットの作成	3
ボックスプロット	6
ボックスプロット・フェンス	6
ボックスプロットのひげの長さ	7
ボックスプロットの外れ値の方法	8
ノッチ付きボックスプロット	9
ボックスプロットの外観オプション	10
バイオリンプロット	14
バイオリンプロットの外観オプション	16
非対称のバイオリンプロット	18
ボックスとバイオリンを異なるように見せる	21
各データポイントを異なるように見せる	22
ボックスプロットとバイオリンプロットの組み合わせ	27

Volume II User's Guide: Part 1 (Graphs #6)

ボックスプロットとバイオリンプロット

Igor Pro マニュアル : II-330 ページ以降をもとに編集

同じ条件を表す複数の測定値がある場合、それらの測定値がどのように分布しているかを知ることが有益です。密集したクラスタ、中央値を中心としたグループ、外れ値を含む緩やかなクラスタなど、さまざまな可能性があります。

単純なポイントの集まりを視覚的に理解することは困難であるため、分布を要約するプロットは有用です。ボックスプロット（箱ひげ図）とバイオリンプロットは、データポイントの分布を要約する 2 つの方法です。

Igor では、ボックスプロットとバイオリンプロットは特殊なグラフィックです。

トレースの各「ポイント」はデータセット全体を表します。

データは個々のウェーブ、各データセットに 1 つのウェーブ、に保存されます。

この場合、トレースの入力は、ウェーブのリストとなります。

あるいは、各データセットが単一の 2 次元マトリックスウェーブの列であるかもしれません。

通常の XY グラフのトレースは、Y データを含むウェーブをもとに名前が付けられます。

したがって、Y データが「wave0」というウェーブにある場合、トレースも「wave0」と名前が付けられます。

しかし、ボックスプロットやバイオリンプロットは、複数のウェーブから得られたデータを表すことがあります。

デフォルトでは、ボックスプロットやバイオリンプロットのトレースは、カスタムのトレース名を指定しない限り、ウェーブリストの最初のウェーブの名前が付けられます。

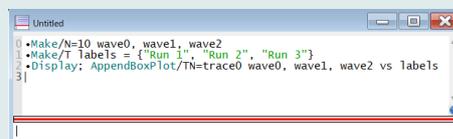
ウェーブの集合の性質を表すカスタムのトレース名を付けることを推奨します。

ボックスプロットとバイオリンプロットの文法

1. 文法は次のようになります。

新しい Experiment で次のコマンドを入力して、実行します。

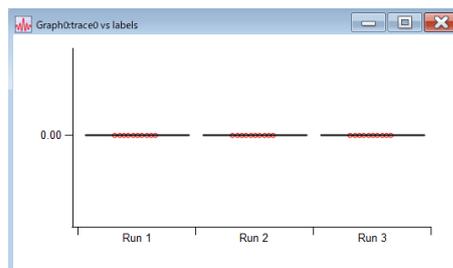
```
Make/N=10 wave0, wave1, wave2
Make/T labels = {"Run 1", "Run 2", "Run 3"}
Display; AppendBoxPlot/TN=trace0 wave0, wave1, wave2 vs
labels
```



2. trace0 という名前の 1 つのトレースを持つボックスプロットが作成されます。

トレースは、wave0、wave1、wave2 という 3 つのデータセットで構成されています。

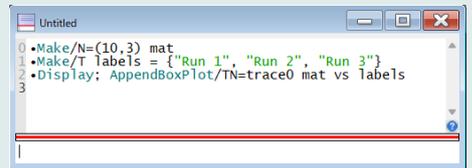
TN=/trace0 を省略すると、トレースのデフォルト名は wave0 になります。



3 つの 1D ウェーブの代わりに、2D の 3 列 2D ウェーブを使っても同じ結果を得ることができます。

1. 新しい Experiment で次のコマンドを入力して、実行します。

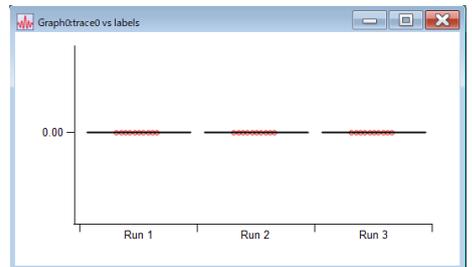
```
Make/N=(10,3) mat  
Make/T labels = {"Run 1", "Run 2", "Run 3"}  
Display; AppendBoxPlot/TN=trace0 mat vs labels
```



2. trace0 という名前の1つのトレースを持つボックスプロットが作成されます。

トレースは、mat[0]、mat[1]、mat[2] という3つのデータセットで構成されています。

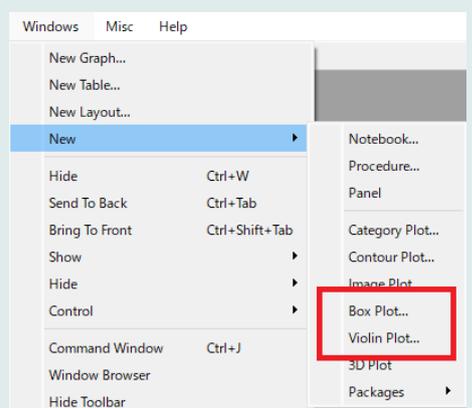
TN=/trace0 を省略すると、トレースのデフォルト名は mat になります。



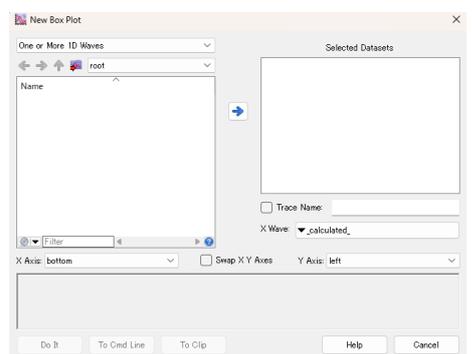
ボックスプロットとバイオリンプロットの作成

1. ボックスプロットを作成するには、Windows → New → Box Plot を選択します。

バイオリンプロットを作成するには、Windows → New → Violin Plot を選択します。



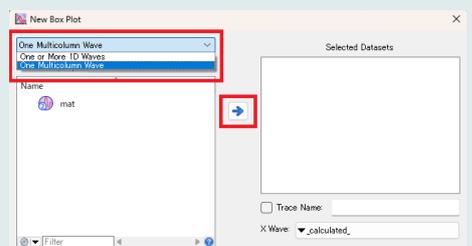
2. これにより、データセットの選択や、その他のパラメーターの設定を行うためのダイアログが表示されます。



3. プロットに使うデータセットを左のリストから選択し、矢印ボタンをクリックして右のリストに移します。

初期状態では、ダイアログには 1D ウェーブのみが表示され、トレースに表示するボックスまたはバイオリンごとに1つのウェーブを選択する必要があります。

データが 2D ウェーブの場合は、左のリストの上にあるポップアップメニューから One Multicolumn Wave を選択します。



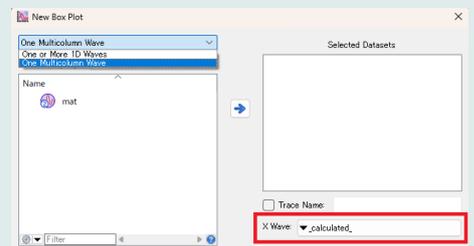
4. 1D ウェーブを使っている場合、ウェーブを右のリストに移した後、上下にドラッグしてウェーブを並べ替えることができます。リスト内のウェーブの順序は、トレース内のプロットの順序を決定します。

複数列のウェーブを使っている場合、その順序はウェーブ内の列の順序によって設定されます。

5. また、X ウェーブを選択する必要があるかもしれません。
calculated を選択すると、X 軸に沿ったプロット的位置は Igor が計算します。

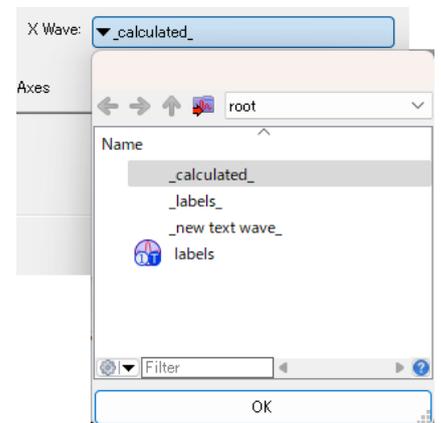
1D ウェーブのリストの場合、プロットは 0、1、2、… の位置に配置されます。

データセットが複数列ウェーブである場合、_calculated_ を選択すると、プロットはウェーブの Y スケーリング、つまり列の大きさのインデックスのスケール値に従って配置されます。

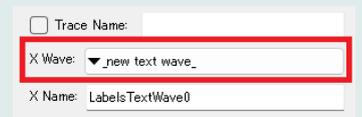


6. X Wave メニューには、数値ウェーブとテキストウェーブの両方が含まれています。
数値ウェーブを選択すると、各プロットを X 軸上の任意のポイントに配置することができます。

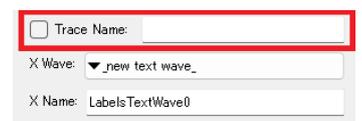
テキストウェーブを選択すると、カテゴリ X 軸が表示されます (マニュアル II-355 Category Plots を参照)。
X Wave メニューに表示されるウェーブは各データセットに1つのポイントがあるものに限られます。



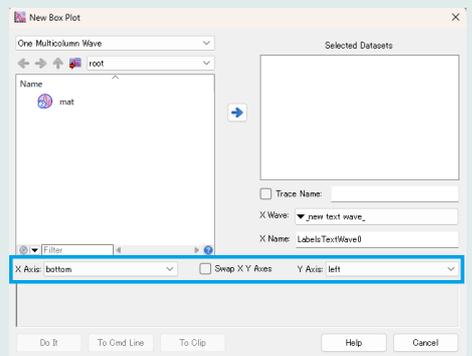
7. New Box Plot ダイアログまたは New Violin Plot ダイアログでも、新しいテキストウェーブを作ることができます。
X Wave メニューから「_new text wave_」を選択すると、ダイアログが適切な長さの新しいテキストウェーブを作成するコマンドを生成し、プレースホルダーテキストで埋め、編集用テーブルに表示します。
その結果、新しいテキストウェーブを使ったカテゴリ X 軸が作成されます。



8. Trace Name チェックボックスをオンにして、編集ボックスに名前を入力すると、新しいトレースにカスタム名前を付けることができます。
デフォルトのトレース名は最初のウェーブの名前をもとにしていてわかりにくいので、1D ウェーブのリストを使う場合はカスタム名が特に便利です。

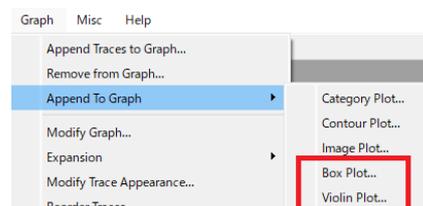


9. X Axis、Y Axis、および Swap XY Axes のコントロールは、New Graph ダイアログ（マニュアル II-277 Creating Graphs を参照）と同じように機能します。



10. グラフには、複数のボックスプロットまたはバイオリンプロットのトレースを保持することができ、2つを混在させることもできます。

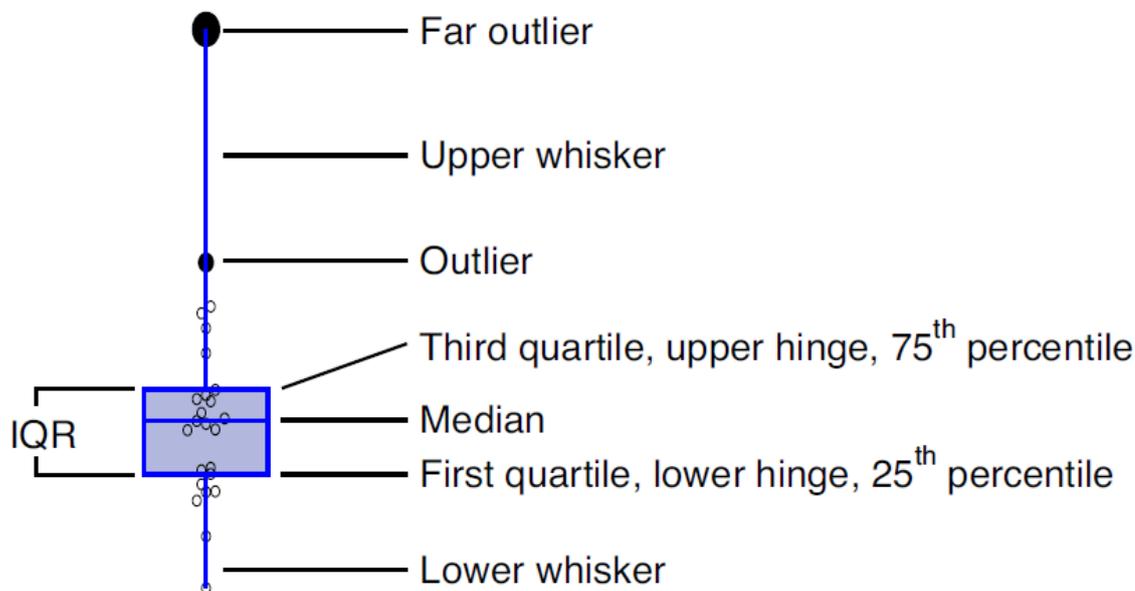
別のボックスプロットまたはバイオリンプロットを追加するには、メニュー Graph → Append to Graph → Box Plot または Graph → Append to Graph → Violin Plot を選択します。



ボックスプロット

ボックスプロット、またはボックス・ウィスカープロットは、データポイントの分布をわかりやすく表示するために、John W. Tukey によって考案されました（マニュアル II-337 Box Plot Reference を参照）。

ボックスプロットはいくつかの部分を持ちます。



ボックスの下辺と上辺は、データセットの第1四分位値と第3四分位値に位置し、箱を横切る線は中央値を表しています。

したがって、ボックスは分布の幅を示し、中央の線は分布の中心位置を示します。

ひげ（ウィスカー）は、分布の対称性やテールの長さなど、データ分布の幅に関するより多くの情報を表します。デフォルトでは、極端なデータポイントまでひげを描画しますが、ひげの長さには8つの異なるオプションがあります。

ボックスの幅は意味を持ちませんが、コントロールすることで見栄えの良い表示にすることができます。

幅は分数で表現でき、プロット領域の幅に対する分数となります。

幅が1より大きい場合、ポイント単位の絶対幅として扱われます。

デフォルトでは、幅は $1/(2*N)$ です。ここで、N はトレースに含まれるデータセットまたはボックスプロットの数です。

分布の中心から離れたとき、実際のデータポイントを表示することが一般に行われます。

Tukey は「外れ値」と「極端な外れ値」を定義しています。

Igor では、すべてのデータポイント、外れ値のみ、極端な外れ値のみ、を表示することができます。

また、データポイントの各カテゴリに対して、異なるマーカー、サイズ、色を選択することもできます。

ボックスプロット・フェンス

一部のオプションでは、Tukey の「フェンス」を使って、ひげの長さや、外れ値および極端な外れ値を定義する境界線を定義します。

Tukey は、25 パーセンタイルと 75 パーセンタイルを指す用語として「ヒンジ (hinge)」も使っています。

内側のフェンスは次のように定義されます：

$$\text{inner fence} = \text{upper hinge} + 1.5 \cdot \text{IQR} \text{ and } \text{lower hinge} - 1.5 \cdot \text{IQR}$$

外側のフェンスは次のように定義されます：

$$\text{outer fence} = \text{upper hinge} + 3 \cdot \text{IQR} \text{ and } \text{lower hinge} - 3 \cdot \text{IQR}$$

ボックスプロットのひげの長さ

図は、ひげの長さを定義するオプションを示しています。

Tukey の定義を参考までにフェンスで示し、外れ値の手法（マニュアル II-334 Box Plot Outlier Methods を参照）はオプション 0 です。

オプション 0：最小と最大のデータポイント（デフォルト）

オプション 1：内側のフェンス

オプション 2：「隣接するポイント」とは、Tukey が内側のフェンス内の最も極端なデータポイントと定義しているもの。つまり、Tukey の定義に従うなら、外れ値ではない最も極端なデータポイント

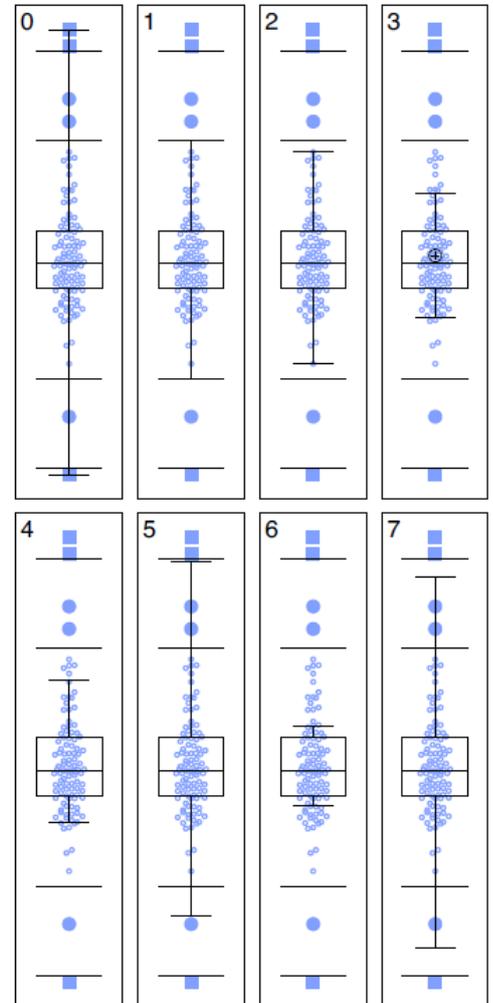
オプション 3：データの平均値からの 1 標準偏差。プラス記号のついた白丸は平均値

オプション 4：9 パーセンタイルと 91 パーセンタイル

オプション 5：2 パーセンタイルと 98 パーセンタイル

オプション 6：任意のパーセンタイル。ここでは、両端を 20 パーセンタイルと 80 パーセンタイルとし、他のオプションとは異なるように設定

オプション 7：平均値から 1 標準偏差を任意の係数で乗算した値。この場合、係数は 3



データが正規分布している場合、2 パーセンタイル、9 パーセンタイル、25 パーセンタイル、50 パーセンタイル、75 パーセンタイル、91 パーセンタイル、98 パーセンタイルは等間隔になっているはずですが。

例では、すべてのデータポイントが表示されていますが、通常は外れ値と極端な外れ値のみを表示します。

データポイントは「ジッター（左右の揺れ）」で表示されます。

重なり合うデータポイントは水平方向にオフセットされ、各データポイントが見えるようになっています。

ボックス幅の小数単位でジッター量を指定することで、オフセットの最大をコントロールできます。

指定した幅が、ポイントを分離するには不十分な場合、必要に応じてポイントが重なります。幅が必要以上に大きい場合、必要なオフセットのみが適用されます。

通常の使用では、フェンスは表示されません。

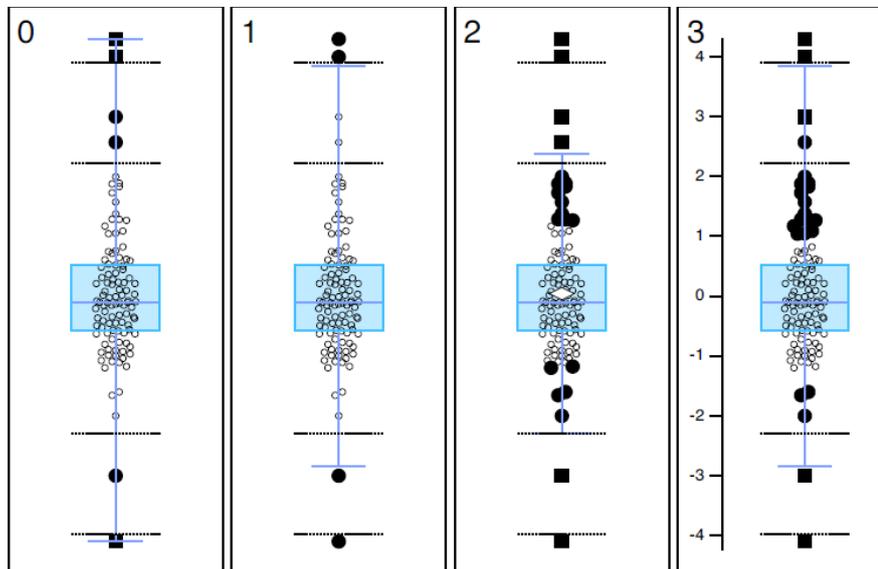
ボックスプロットの外れ値の方法

デフォルトでは、Igor は Tukey に従って、フェンスに基づいて外れ値と極端な外れ値を定義し、「外れ値」は内側のフェンスの外側のポイント、「極端な外れ値」は外側のフェンスを越えたポイントとしています。

通常は表示されませんが、プロット上にフェンスを含めるように設定するには、コマンド `ModifyBoxPlot showFences=1` を使います。

外れ値は塗りつぶした円で表示され、極端な外れ値は塗りつぶした四角で表示されます。

Igor は、どのデータポイントが外れ値および極端な外れ値であるかをコントロールするための4つのオプションを提供しています。



オプション 0 : Tukey の定義。外れ値とは、内側のフェンスを越えたデータポイントを指し、極端な外れ値とは外側のフェンスを越えたデータポイントを指す

オプション 1 : ひげの端を越える点はすべて外れ値。極端な外れ値はなし。上図では、ひげはオプション 6 (前のセクションのオプション)、2 パーセンタイルと 98 パーセンタイルで設定

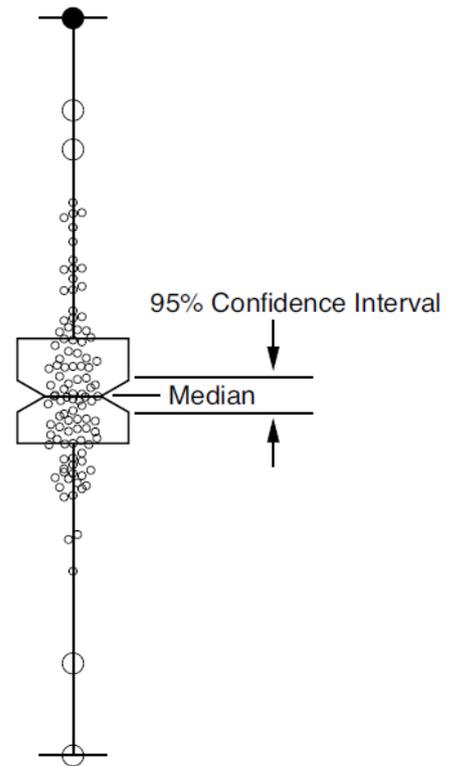
オプション 2 : 外れ値および極端な外れ値とは、平均の標準偏差に任意の係数を乗じた値を超えるポイントを指す。上図では、係数は 1 と 2。ひげの長さはオプション 7 (前のセクションのオプション) に設定され、平均値の標準偏差に任意の係数を乗じた値。係数は 2 に設定。中央の白いひし形は平均値

オプション 3 : 外れ値および極端な外れ値は、4つの任意のデータ値によって決定。上図の場合、値は -2、-1.5、1、2.6

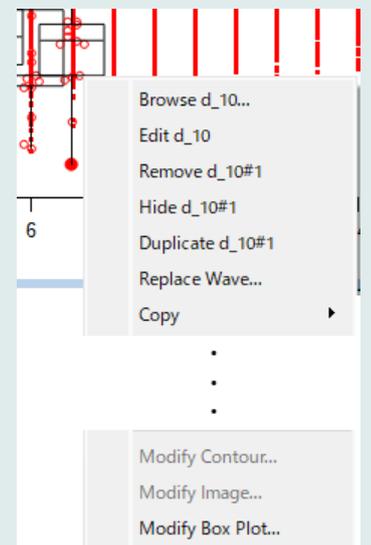
ノッチ付きボックスプロット

ノッチ（切り欠き）付きボックスプロットは、通常表示されるさまざまなパーセンタイルに加えて、中央値の 95% 信頼区間を示します。

ノッチは中央値 $\pm 1.57 * IQR / \sqrt{n}$ で、 n はボックスプロットで表されるデータセット内のデータポイント数です。
ノッチが重なっていない場合、2つの中央値が大きく異なる可能性が高くなります。

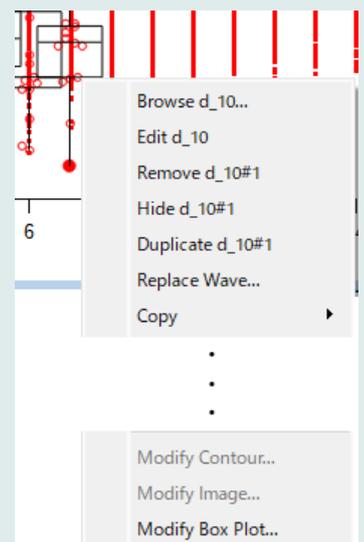


ボックスプロットのトレースを右クリックして、Modify Box Plotを選択すると、ノッチ付きボックスプロットを作ることができます。



ボックスプロットの外観オプション

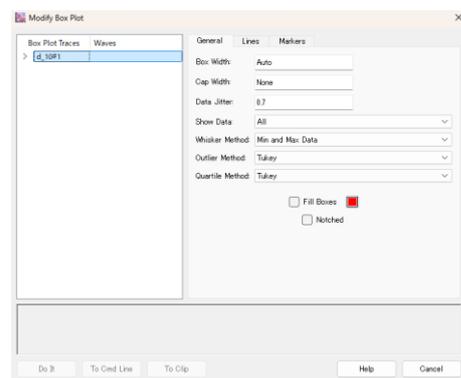
1. ボックスプロットの外観を変更するには、メニュー Graph → Modify Box Plot を選択するか、ボックスプロットのトレースを右クリックしてコンテキストメニューから Modify Box Plot を選択します。



2. 線の色、幅、破線のスタイルを変更できます。

これらの設定は、ボックス、ひげ、中央線、ひげのキャップのコンポーネントごとに設定できます。

私たちはフェンスを付随的な詳細情報として考えています。表示をより理解してもらうためにフェンスを表示することはできませんが、出版目的で使われているのを見たことがありません。したがって、フェンスの線は修正できません。



ひげのキャップの幅をコントロールすることができます。

キャップを省略する（デフォルト）場合は、幅をゼロにします。

幅を分数で設定すると、ボックスの幅に対する分数としてキャップのサイズが設定されます。

これは、グラフのサイズを変更するときにキャップの幅を一定に保つのに便利です。

1 より大きい値は、ポイント単位の絶対的な幅として扱われます。

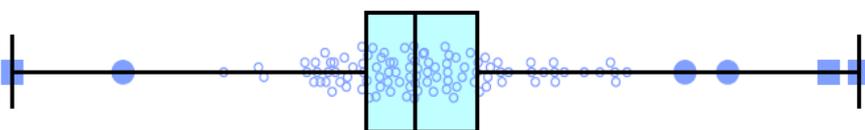
次のプロットは、スペースを節約するために水平方向に描画されています。

メニュー Graph → Modify Graph を選択し、Swap XY チェックボックスをオンにすることで実現できます。

あるいは、1つのボックスプロットのトレースだけを入れ替えるために、New Box Plot ダイアログ、または Append Box Plot ダイアログで Swap XY Axes チェックボックスをオンにすることができます。

ボックスは色で塗りつぶすことができます。

ボックスの塗りつぶしはマーカーより先に描画されるため、マーカーにどのような選択をしても、塗りつぶしによって上書きされることはありません。

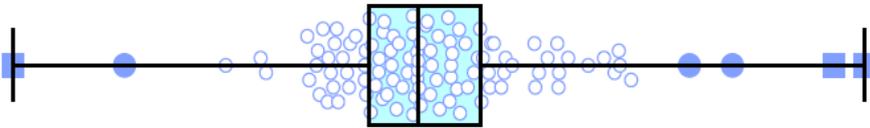


外れ値ではないデータ、外れ値、極端な外れ値には、それぞれ異なるマーカーを選択できます。

上の図では、通常のデータポイントは中空の円で、外れ値は塗りつぶされた円で描画されていて、外れ値は通常のデ

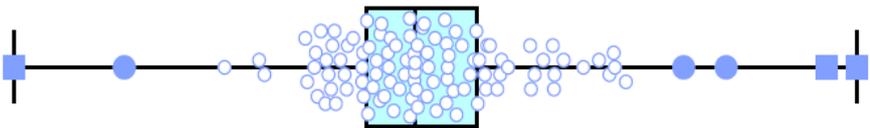
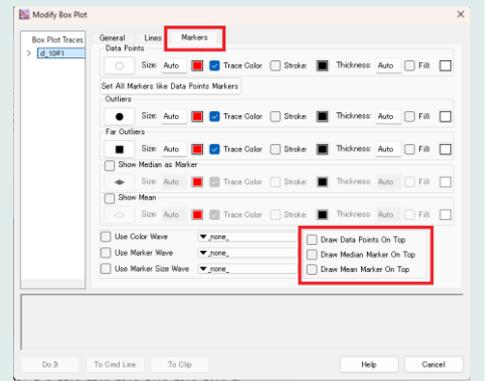
ータポイントよりも大きく描画されて強調されています。
極端な外れ値は、塗りつぶされた四角のマーカで描画されています。

中空のマーカを使って、マーカに塗りつぶし色を選択することで、特殊効果を得ることができます。
例えば、次の図では、通常のデータポイントに中空の円マーカを使い、塗りつぶし色を有効にして白に設定しています。

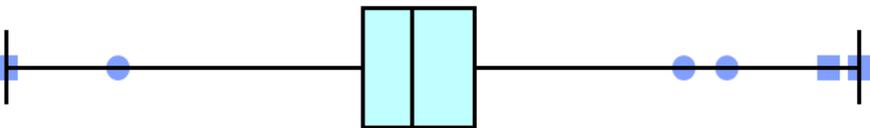


ボックス、中央値、ひげを描く線はマーカの上に表示されるため、マーカが線を見えにくくすることはありません。

このデータセットでは、線が不明瞭になりすぎる可能性があります
が、マーカを上に表示したい場合もあります。
Markers タブの Draw Data Points on Top チェックボックスを
オンにします。
平均マーカと中央値マーカの描画は個別に制御できます。



上記のようにすべてのデータを表示するか、外れ値ではないデータを表示せずに外れ値と極端な外れ値のみを表示するか、極端な外れ値のみを表示するかを選択できます。
次のプロットは、外れ値と極端な外れ値のみを示しています。



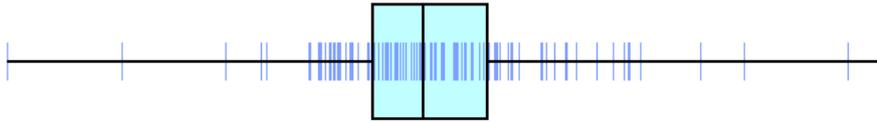
これまで示したプロットでは、データポイントにジッターが適用されています。
つまり、マーカが重なり合う場合は、すべてのマーカが見えるように横方向のオフセットが適用されます。
ジッターの幅を指定することで、横方向の最大オフセットを設定することができます。
オフセットが小さすぎてマーカを完全に分離できない場合、ある程度重なり合います。
特定のマーカにオフセットが不要な場合は、オフセットされません。

データポイントが「ラグプロット（軸上の線として表現するプロット）」として表示されるボックスプロットを目にすることがあります。
つまり、各データポイントは細い線で表されます。
データポイントの数が多すぎない限り、これもすべてのデータポイントを確認できる方法の1つです。

これを行うには、ラインマーカーを使い、ジッターをゼロに設定します。

次のプロットでは、通常値、外れ値、極端な外れ値のマーカーはすべて同じ縦線マーカー（マーカー 10）で、サイズは 8 に設定されています。

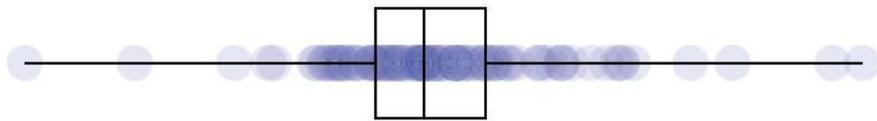
キャップはゼロに設定されていて、極端なデータポイント（両端）はキャップのように見えます。



透明度の設定ができる大きなマーカーも、データの密度感を表現することができます。

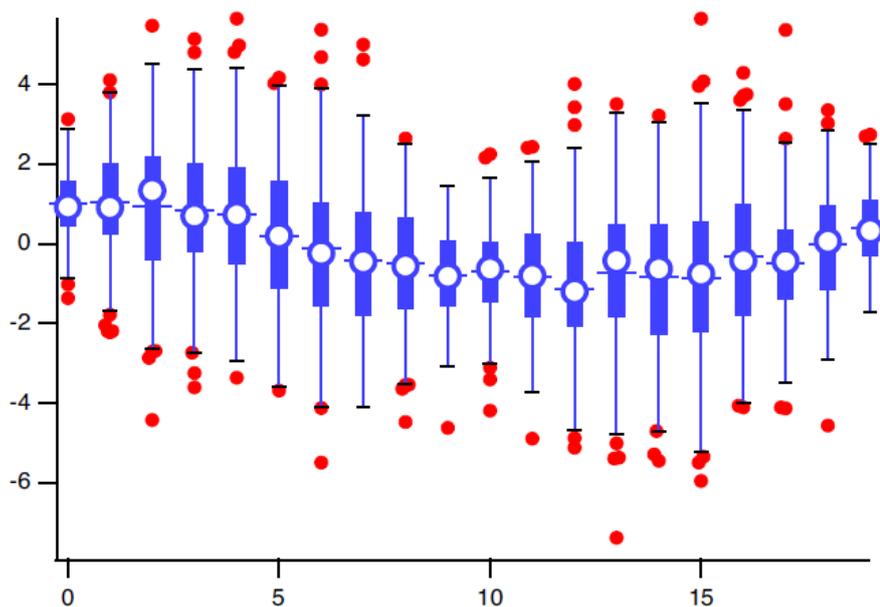
次のプロットでは、アルファ値を 0.1 に設定した青色のソリッドの円マーカー（マーカー 19）を使っています。

透明マーカーでは背景が透けて見え、色に影響を与えるため、ボックスの塗りつぶし色はオフにしています。



データセットやデータポイントの数が多すぎて圧倒されることがあります。

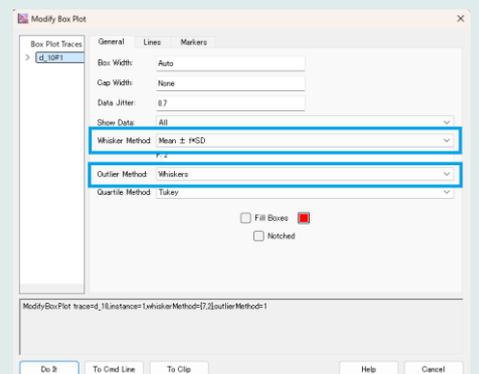
次のような、よりコンパクトな表示にするための代替案もあります。



Modify Box Plot ダイアログで、Whisker Method は Mean \pm f*SD に設定され、係数は 2 に設定されているため、ひげは平均値の 2 標準偏差の幅を示します。

Outlier Method は Whiskers に設定されているため、ひげの端を越えるデータポイントのみが表示されます。

これにより、プロット上のデータポイントの数が大幅に削減されます。



データセットが多数あるため、ボックスはかなり狭く、塗りつぶされてしっかりとしたボックスになっています。見えにくい中央線ではなく、中央値は白く塗りつぶされた円のマーカで表示されています。白のマーカの塗りつぶしの上にボックスの輪郭が表示されないように、Modify Box Plot ダイアログの Markers タブで、Draw Median Marker On Top チェックボックスをオンにします。

最後に、平均値は水平の棒グラフマーカとして表示されます。

マーカのサイズは、中央の円マーカの外側に表示されるように十分に大きく設定されています。

バイオリンプロット

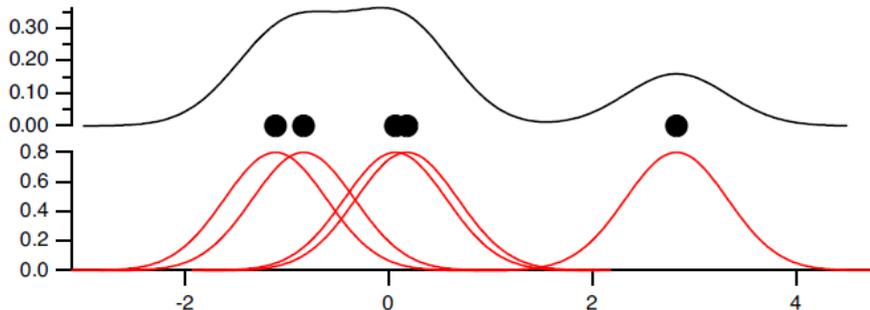
バイオリンプロット（別名：ビーン[豆]プロット）は、データの分布を要約する方法です。

バイオリンプロットは、カーネル密度推定（kernel density estimate [KDE]）を使って、データセットの分布を表示します。

KDE は、データポイントごとに1つのカーネル関数となるいくつかのカーネル関数を合計することで、基礎となるデータ分布の滑らかな推定値を作ります。

次に、合計された曲線は、データセットの確率分布関数の推定値となるように、面積が 1.0 に正規化されます。

例えば、次に示した黒い点のような、ガウス分布から5つの点が描画されたとします。



各ポイントに対してガウス曲線（赤い曲線）を生成し、その曲線を合計して、1つの領域として正規化します（黒い曲線）。

このグラフでは、赤い曲線の標準偏差を 0.5 と、任意に選択しています。

これは、KDE 曲線やバイオリンプロットを計算するときの帯域幅と呼ばれます。

これで、データポイントが描画された基礎となる分布を示すことができる滑らかな曲線が得られました。

使用したカーネルの帯域幅が小さすぎるため、ポイントが少ないことで正当化できない可能性もあります。

カーネル関数の選択、この例ではガウス関数、およびカーネルの幅は、ある程度任意です。

ガウスカーネルはある意味「滑らか」であり、ほとんどのデータはガウス分布に従うという先入観を反映しています。

もちろん、他にも選択肢はあります。

StatsKDE コマンドを使うと、自分でデータセットの KDE を計算することができますが、そこからバイオリンプロットを作るのはかなり面倒です。

マニュアル II-330 Creating Box Plots and Violin Plots で説明している手順を実行することで、バイオリンプロットトレースを作るときの多くの作業を行ってくれます。

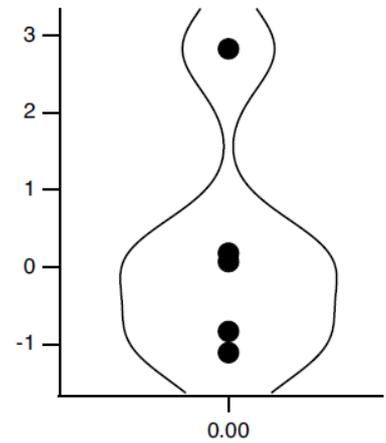
バイオリンプロットでは、一般的に曲線は垂直方向にプロットされ、中央線に対称となり、バイオリンまたは各豆のところにこぶのあるインゲン豆のさやのようなプロットが作成されます。

右図は上記の5つのポイントのバイオリンプロットで、データポイントを表示し、黒で色を付けるように修正したものです。

このグラフでは、曲線の幅は相対的な場合を除いて意味を持ちません。

中心線は推定確率密度がゼロであることを示しています。

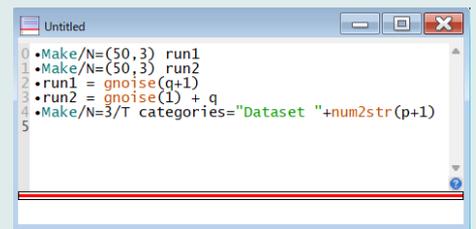
これはトレースに含まれる唯一のデータセットであり、カテゴリプロットではないため、X 軸のゼロの位置に配置されます。



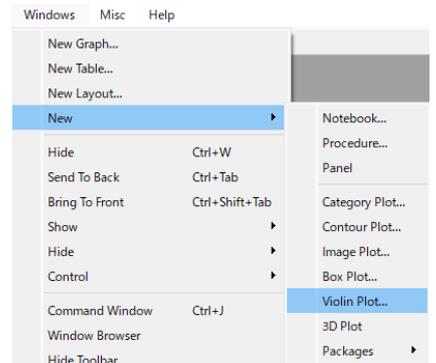
以下では、2つのフェイクのデータセットを作成して、表現を確認します。

1. 新しい Experiment を作成し、コマンドウィンドウで次を実行します。

```
Make/N=(50,3) run1 // 50 ポイント/データセット、3 データセット
Make/N=(50,3) run2 // 50 ポイントの別の 3 データセット
run1 = gnoise(q+1) // 標準偏差付きガウス分布データ
run2 = gnoise(1) + q // 一定幅と場所を持つガウス分布データ
Make/N=3/T categories="Dataset "+num2str(p+1)
// カテゴリプロット用テキストウェーブ
```

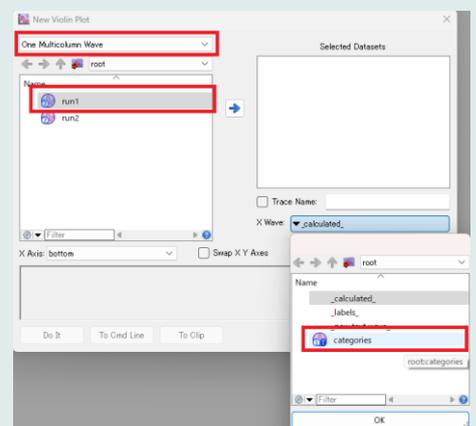


2. メニュー Windows → New → Violin Plot を選択します。



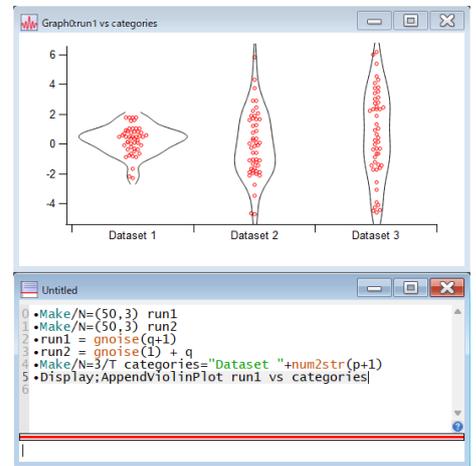
3. プルダウンメニューから One Multicolumn Wave を選択し、run1 を選択して、右の列に追加し、X Wave のプルダウンメニューから categories を選択します。

Do It をクリックします。



4. これにより、次のコマンドが生成され、プロットが表示されます。

```
Display;AppendViolinPlot run1 vs categories
```

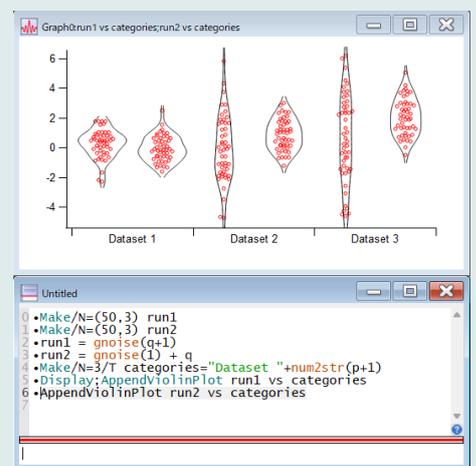


5. メニュー Graph → Append To Graph → Violin Plot を選択し、One Multicolumn Wave を選択し、run2 を選択して、右の列に追加し、X Wave のプルダウンメニューから categories を選択します。

Do It をクリックします。

次のコマンドが生成され、プロットは右のようになります。

```
AppendViolinPlot run2 vs categories
```



バイオリンプロットの外観オプション

バイオリンプロットの外観を変更するには、メニュー Graph → Modify Violin Plot を選択するか、バイオリンプロットを右クリックして Modify Violin Plot を選択します。

バイオリンプロットでは、6つのカーネル形状から選択できます。

デフォルトはガウスクーネルで、他の形状はあまり使われることはないと思います。

使用する最適な帯域幅を自動的に推定する方法は3つあります。

自動帯域幅の結果が気に入らない場合は、各データセットに個別の帯域幅を設定して、独自の帯域幅を設定することができます。

自動の推定はガウスクーネルを前提としています。

バイオリンプロットの各プロットは、ボックスプロットのボックス幅に相当する水平方向のスペースを使用します。ボックス幅はボックスプロットと同じ方法で設定できます。

分数はプロット領域の分数に相当する幅を指定し、1より大きい値はポイント単位の絶対サイズです。

X軸がカテゴリ軸の場合、ボックスの幅を指定し、ユーザーが選択した内容を無効にします。

複数のデータセットを含むバイオリントレースを作成すると、すべての曲線は KDE 曲線の中で最も大きなピークに合わせて正規化され、バイオリンプロットの1つだけがボックス全体の幅を占めることとなります。

これは前のセクションの run1 トレース（各カテゴリの左のプロット）の3番目のデータセットは、より広い分布を持つデータセットを表していて、1つの領域を実現するには、振幅をより小さくする必要があることを意味します。他のカテゴリの、短い、太い分布と比較してみましょう。

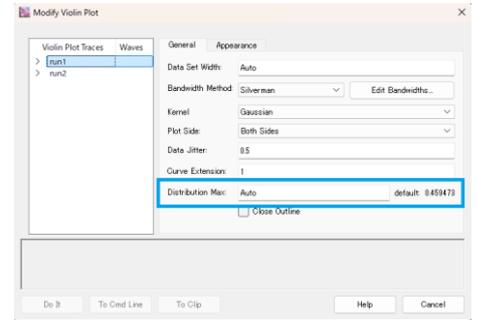
この例のように、バイオリンプロットがセットになっている場合、デフォルトでは、各トレースで最も幅の広いプロットがボックス全体を埋めます。

2つのトレースは別々に正規化されます。

両方のトレースのすべてのプロットに同じ正規化を使いたい場合は、すべての曲線の相対振幅が同じ相対値を表すように、正規化を自分で設定することができます。

これは Modify Violin Plot ダイアログの General タブにある Distribution Max 設定を使って行います。

Igor によって計算された値は、編集ボックスの隣に表示されます。



マーカー付きのデータを表示することは一般的に行われます。

デフォルトでは、データポイントは通常のトレースのデフォルトマーカーの半分のサイズで描画された中空の円で表示されます。

データポイント用のマーカーを選択でき、さらに必要に応じて平均値と中央値用のマーカーも選択できます。

バイオリンプロットでは外れ値を区別することはできません。

データポイントのマーカーにジッター（左右の揺れ）を適用して、それらが重ならないようにすることができます。

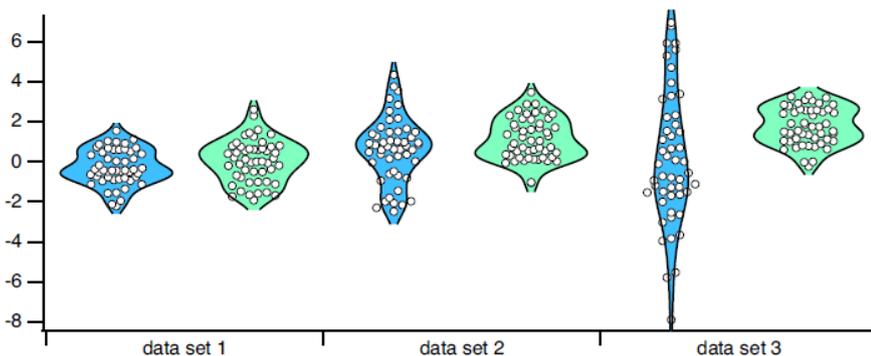
データのラグプロット（軸上の線として表現するプロット）を作成するには、適切なラインマーカーを使って、ジッターを 0 に設定します。

ほとんどのバイオリンプロットは縦軸でプロットされているため、水平方向のマーカーが適切です。

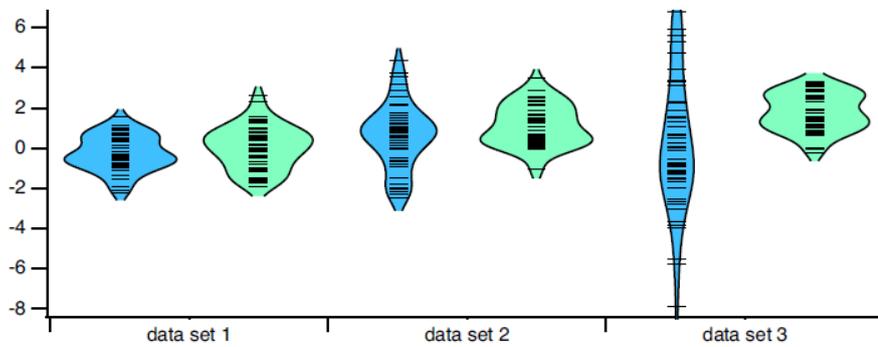
KDE の曲線の間スペースは色で埋めることができます。

塗りつぶしの色はマーカーの後ろに描画されるため、マーカーの色、マーカーの線の色、マーカーの塗りつぶし色で特殊効果を得ることができます。

下の図は、塗りつぶし色と、白で塗りつぶした中空の円マーカーで再加工したプロットです。



次はラグプロットです。



一部の曲線が少し途切れているように見えるかもしれません。

特に、data set 3 の右側のトレースがそうです。

デフォルトでは、KDE 曲線を最後のデータポイントを越えた 1 帯域幅先にプロットします。

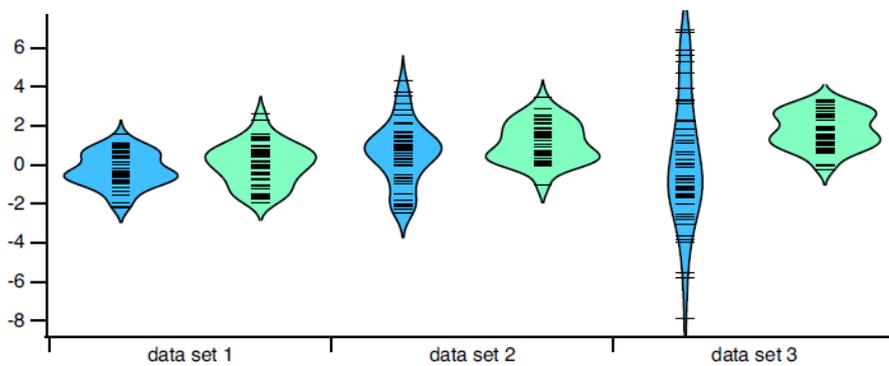
この場合、曲線をもう少し伸ばしてもよいように思われます。

それを達成するには、トレースを右クリックして Modify Violin Plot を選択します。

General タブで Curve Extension 編集ボックスにより好ましい延長量を入力します。

単位はカーネル帯域幅です。

下図では、上図のプロットに 2 を設定しています。



非対称のバイオリンプロット

標準的なバイオリンプロットの一般的なバリエーションは、バイオリンの半分だけ使うものです。

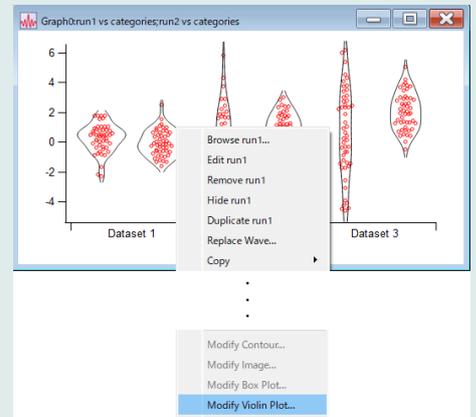
これは通常、2つの条件または2つのランを比較するために使われます。

1つの条件が左半分を提供し、他方が右半分を提供します。

このセクションの最初に作ったバイオリンプロットをもとに非対称プロットを作る手順を確認します。

1. プロット内のトレースの1つを右クリックし、コンテキストメニューから **Modify Violin Plot** を選択します。

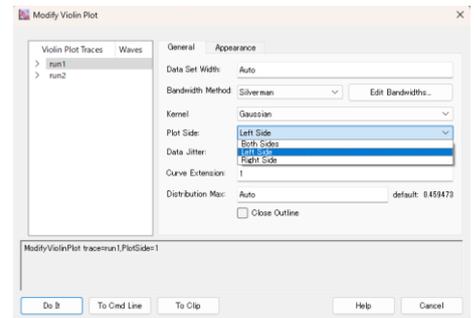
※注：マニュアルでは、ラグプロットの図で説明されています。



2. 左半分に表示したいトレースを選択します。

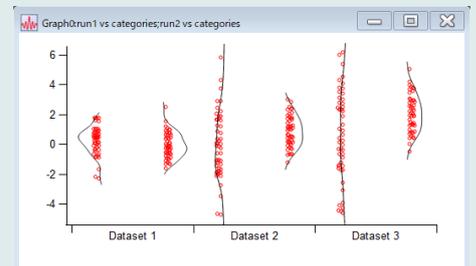
例えば、run 1 トレースを選択します。

General タブで Plot Side メニューから **Left Side** を選択します。



3. run2 を選択し、Plot Side メニューで **Right Side** を選択します。

Do It をクリックすると、右のようなプロットが作成されます。



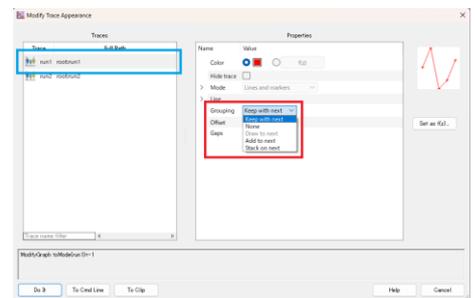
4. X 軸はカテゴリ軸であるため、各プロットはカテゴリ内のそれぞれのスロットを占めます。

これらをまとめるには、run1 トレースをダブルクリックして、**Modify Trace Appearance** ダイアログを表示します。

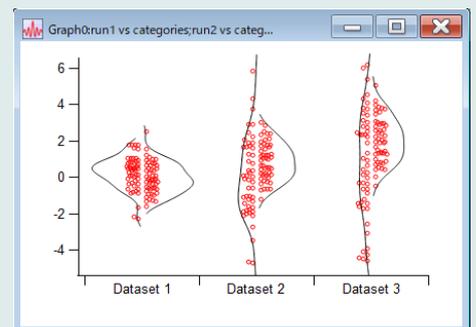
トレースリストの最初のトレースが選択されていることを確認します。

Grouping メニューで **Keep with next** を選択します。

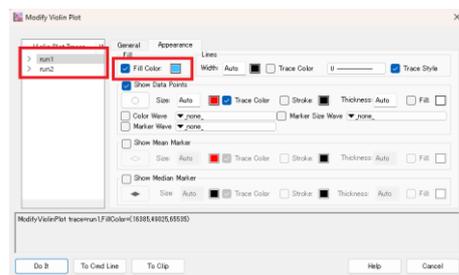
X 軸が数値軸の場合は、Grouping モードの設定は不要です。



5. Do It をクリックすると、右のようになります。

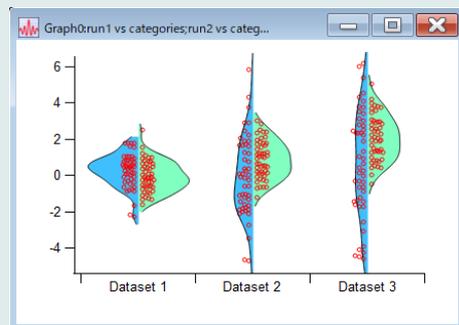


6. オリジナルのマニュアルに合うように、色を付けておきます。
Modify Violin Plot ダイアログで、run1 に対して Fill Color で青を、run2 に対しては緑を選択して、Do It をクリックします。

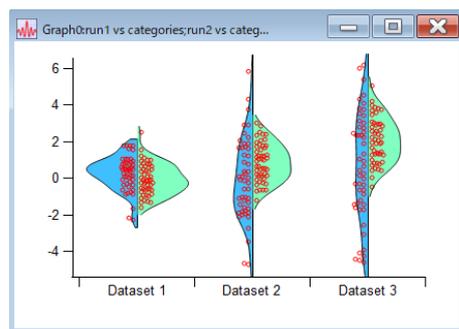


7. プロットが着色されます。

2つのバイオリンプロットのトレースは、X Wave で設定された数値、または `_calculated_` を選択した場合は、0、1、2 などの数値でプロットを描画します。



8. 最後に、色で塗りつぶされた領域全体を囲む線があるほうがプロットが良くなると思われるかもしれません。
その時は、Modify Violin Plot ダイアログの General タブで Close Outline チェックボックスをオンにします。



ボックスとバイオリンを異なるように見せる

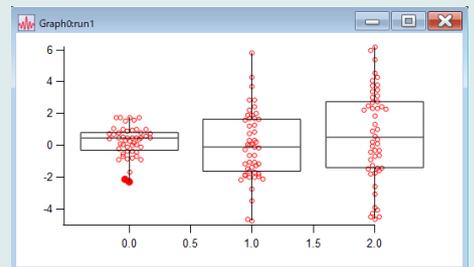
トレース内のボックス、またはバイオリンの1つを、トレース内の他のデータセットと区別するために、異なる外観にしたい場合があるかもしれません。Modify Box Plot ダイアログと Modify Violin Plot ダイアログでは、トレースから単一のデータセットを選択することで、これをサポートしています。

1. ここまでで使用してきたデータを使って手順を確認します。
(マニュアルとは表示が少し異なるので注意)

メニュー **Windows** → **New** → **Box Plot** を選択し、新しいグラフを作ります。

New Box Plot ダイアログで、左上のプルダウンメニューから **One Multicolumn Wave** を選択します。

下のリストから **run1** を選択し、右に移動し、**Do It** をクリックします。

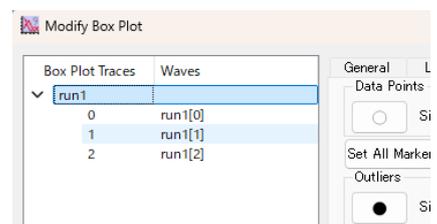


2. メニュー **Graph** → **Modify Box Plot** を選択します。

図は、3つのウェーブのリストで作られたボックスプロットのトレースを含む **Modify Box Plot** ダイアログを示しています。

リストの最初のウェーブは **run1** なので、トレースの名前は「run1」です。

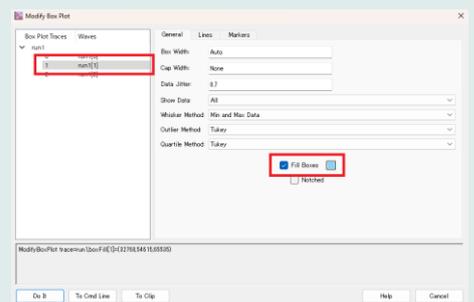
run1 トレースの左の三角アイコンをクリックすると、3つのデータセットのリストが開き、各ウェーブが表示されます。



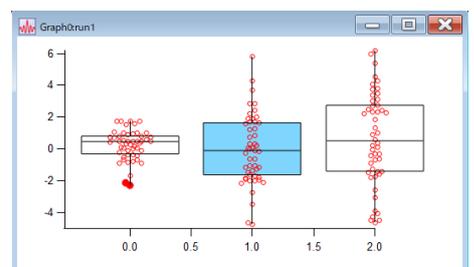
3. 一番上の **run1** が選択されている場合は、このダイアログで変更した内容はトレース内のすべてのデータセットに適用されます。

ここでは、**wave1** に対応する2番目のデータセット **run1[1]** を選択し、ダイアログの変更はこのデータセットにのみ適用します。結果として、2番目のボックスプロットにのみ適用されます。

General タブの **Fill Boxes** チェックボックスオンにし、水色で塗りつぶします。



4. **Do It** をクリックすると、プロットに適用され、2番目のボックスにのみ色が付けられます。



これらの設定を行う順序は重要です。

全体的な設定を変更すると、それまでに設定された個々のデータセットに対応する値はすべて、全体的な値にリセットされてしまいます。

オリジナルのマニュアルの画像でいうと、ボックスプロットの全体を水色で塗り、2番目のデータセットを薄緑色で表示するには、まずトレース全体に水色を設定し、次に2番目のトレースに薄緑色を設定します。

この2つの変更の間でダイアログを閉じる必要はありません。

各データポイントを異なるように見せる

単一のデータセット内のデータポイントが異なるソースから取得されていたり、異なる条件を表している場合があります。それをグラフで表示したい場合があります。

補助ウェーブを使うと、各データポイントに対してマーカーの色、サイズ、スタイルを設定することができます。これらの設定のコントロールは、Modify Box Plot ダイアログの Markers タブに表示されます。

例えば、このテーブルでは、最初の列にフェイクのデータセットが表示され、各データポイントのマーカーの色を指定する3列のウェーブが示されています。

DifferentMarke	BoxPlotColors[BoxPlotColors[BoxPlotColors[
Data Set 1	0	1	2
1.12358	32768	40777	65535
-0.670571	32768	40777	65535
0.696167	32768	40777	65535
-0.48165	5545.79	58663.6	17962.9
-0.888855	5546	58664	17963
-1.81002	5546	58664	17963
0.883285	5546	58664	17963
-0.854401	65535	43690	0
-0.908264	65535	43690	0
-0.682166	65535	43690	0
-0.979681	65535	43690	0
1.90386	65535	43690	0
1.79543	5546	58664	17963
1.20602	5546	58664	17963
1.28203	32768	40777	65535
2.13408	32768	40777	65535
1.78677	32768	40777	65535
1.04066	65535	43690	0
3.2046	65535	43690	0
2.31871	65535	43690	0

以下、手順を追いかけてみます。

1. 新しい Experiment を作成し、テーブルに上記の一番左の値を入力します。

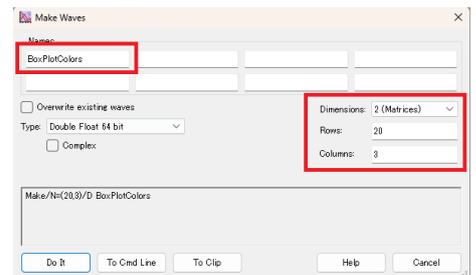
Point	DifferentMarker
0	1.12358
1	-0.670571
2	0.696167
3	-0.48165
4	-0.888855
5	-1.81002
6	0.883285
7	-0.854401
8	-0.908264
9	-0.682166
10	-0.979681
11	1.90386
12	1.79543
13	1.20602
14	1.28203
15	2.13408
16	1.78677
17	1.04066
18	3.2046
19	2.31871
20	

2. 次にカラー指定のウェーブを作成します。

メニュー Data → Make Waves を選択します。

Make Waves ダイアログで、ウェーブ名「BoxPlotColors」を入力し、Dimensions で「2 (Matrices)」を選択して、20 行 3 列と指定します。

Do It をクリックします。



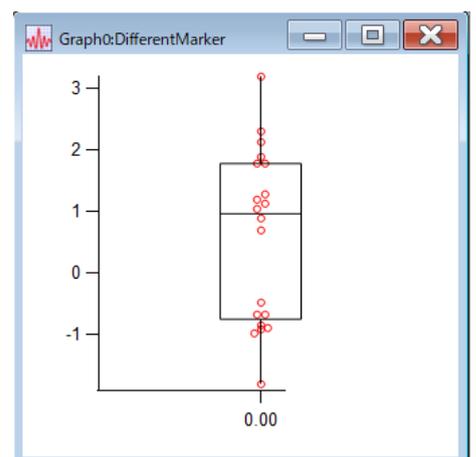
3. BoxPlotColors のテーブルを開き、値を入力します。

Row	BoxPlotColors[[0]]	BoxPlotColors[[1]]	BoxPlotColors[[2]]
0	32768	40777	65535
1	32768	40777	65535
2	32768	40777	65535
3	5545.79	58663.6	17962.9
4	5546	58664	17963
5	5546	58664	17963
6	5546	58664	17963
7	65535	43690	0
8	65535	43690	0
9	65535	43690	0
10	65535	43690	0
11	65535	43690	0
12	5546	58664	17963
13	5546	58664	17963
14	32768	40777	65535
15	32768	40777	65535
16	32768	40777	65535
17	65535	43690	0
18	65535	43690	0
19	65535	43690	0
20	65535	43690	0

4. メニュー Windows → New → Box Plot を選択します。

New Box Plot ダイアログで、One or More 1D Waves を選択し、DifferentMarker を右に移動します。

その他の項目は変えずに Do It をクリックします。

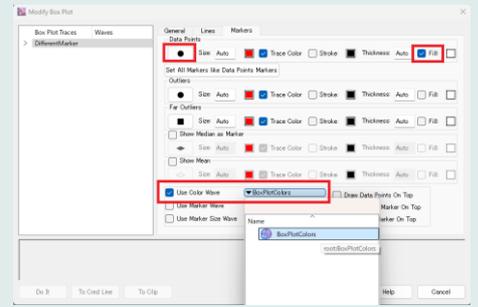


5. ボックスプロットが表示されたら、Modify Box Plot ダイアログを表示します。

Use Color Wave チェックボックスをオンにして、プルダウンメニューから BoxPlotColors ウェーブを選択します。

(マーカが空白でないようにしてください)

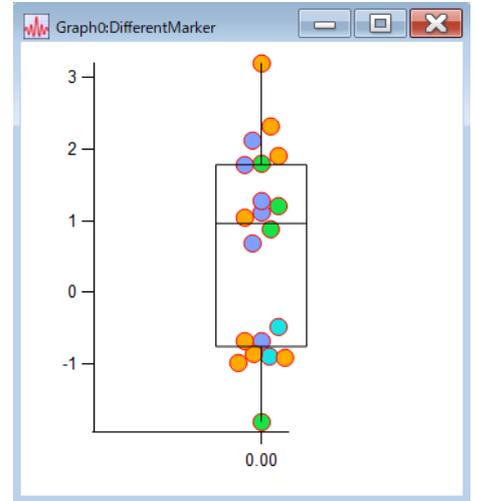
Do It をクリックします。



6. 右のようなグラフに変わります。

(色がわかりやすいように、マーカを大きくしています)

この図のカラーウェーブは、メニュー Data → Packages → Color Wave Editor を選択してアクセスできる Color Wave Editor パッケージを使って作成しています。



7. 同様に、各データポイントにサイズ、またはマーカ番号を指定した 1D ウェーブを使って、各データポイントマーカを異なるサイズ、または異なるマーカスタイルに設定することができます。

テーブルで次のような新しい 1D ウェーブを 2 つ入力します。

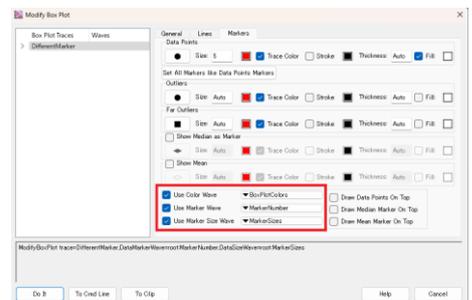
Point	DifferentMarker	MarkerNumber	MarkerSizes
0	1.12358	16	5
1	-0.670571	16	5
2	0.696167	16	5
3	-0.48165	19	5
4	-0.888855	19	7
5	-1.81002	19	7
6	0.883285	19	10
7	-0.854401	18	10
8	-0.908264	18	10
9	-0.682166	18	4
10	-0.979681	18	4
11	1.90386	17	4
12	1.79543	17	4
13	1.20602	60	4
14	1.28203	60	4
15	2.13408	60	5
16	1.78677	60	5
17	1.04066	29	5
18	3.2046	29	6
19	2.31871	29	6
20			

DifferentMarker	MarkerNumber	MarkerSizes
0		
1.12358	16	5
-0.670571	16	5
0.696167	16	5
-0.48165	19	5
-0.888855	19	7
-1.81002	19	7
0.883285	19	10
-0.854401	18	10
-0.908264	18	10
-0.682166	18	4
-0.979681	18	4
1.90386	17	4
1.79543	17	4
1.20602	60	4
1.28203	60	4
2.13408	60	5
1.78677	60	5
1.04066	29	5
3.2046	29	6
2.31871	29	6

8. Modify Box Plot ダイアログを表示します。

Use Color Wave を BoxPlotColors に、Use Marker Wave にチェックを入れ MarkerNumbers を、Use Marker Size Wave にチェックを入れ MarkerSizes を指定します。

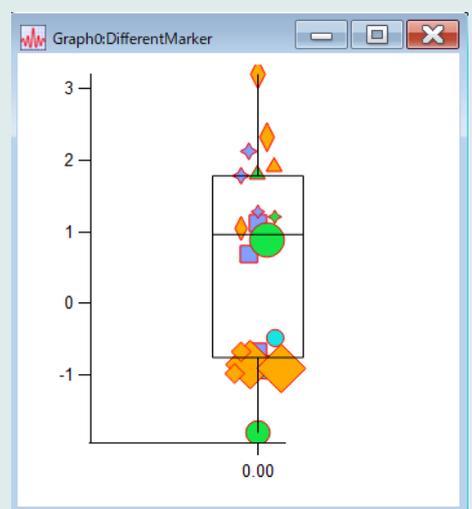
Do It をクリックします。



9. 指定したウェーブのデータを使って、マーカーの形状や大きさが変わりました。

マーカーサイズは 0 から 200 の範囲で指定できます。

マーカーの種類と関連するマーカー番号の一覧は、マニュアル II-291 Markers を参照してください。



これらのウェーブで指定されたマーカーの色、サイズ、スタイルは、他の設定をすべて上書きします。

例えば、ボックスプロットの外れ値として特定のマーカーを選択した場合でも、マーカーウェーブで外れ値のデータポイントにマーカーの設定があると、マーカーはウェーブから取得されます。

マーカーの色、サイズ、スタイルのウェーブはデータセットのウェーブと同じ行数である必要はありません。

余分な行がある場合は、余分な行は無視されます。行数が少ない場合は、マーカーウェーブに値のないデータポイントは、色、サイズ、スタイルは、通常の設定から取得します。

ボックスプロットとバイオリンプロットの組み合わせ

ボックスプロットの1つの問題は、データの真の分布を隠してしまうことです。

データが二峰性の場合でも、中央線とひげのついたボックスのみが表示されます。

一方、バイオリンプロットには、ボックスプロットに含まれるような統計情報が欠けています。

一般的な解決策は、バイオリンプロットの中央にボックスプロットを配置することです。

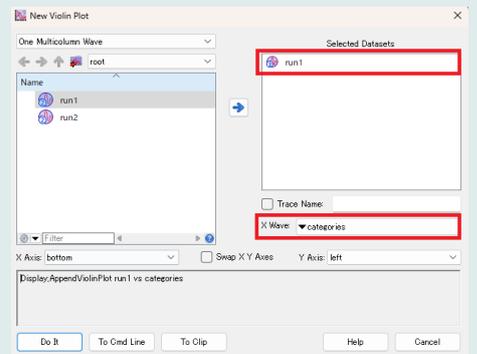
前述の「バイオリンプロット」のセクションのデータを使って、手順を確認します。

コマンドを使ってデータを生成した状態から始めます。

1. メニュー Windows → New → New Violin Plot を選択して、バイオリンプロットを作成します。

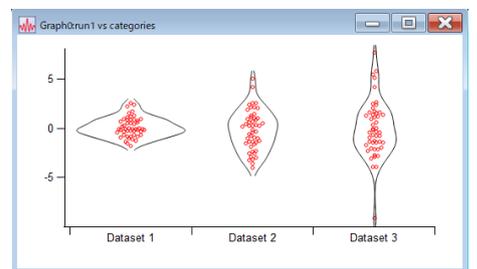
使うデータセットは run1 を使います。

X Wave では categories を選択します。



2. データポイントを表示しないようにし、バイオリンプロットを塗りつぶします。

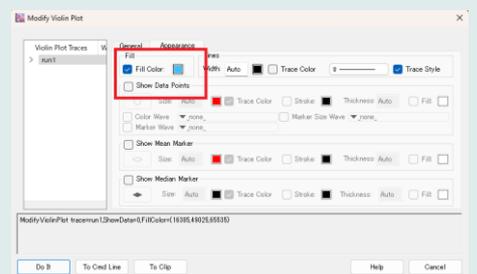
メニュー Graph → Modify Violin Plot を選択します。



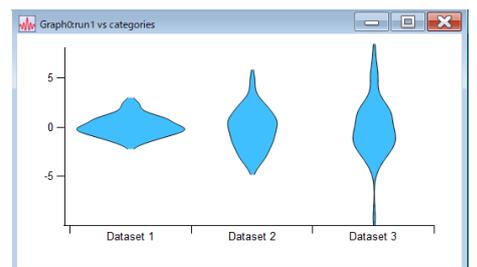
3. Appearance タブで、Fill Color にチェックを入れ、青色を指定します。

Show Data Points チェックボックスをオフにします。

Do It をクリックします。

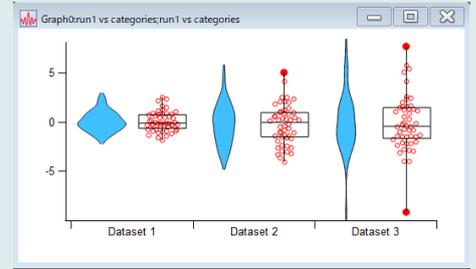


4. バイオリンプロットが右のようになります。



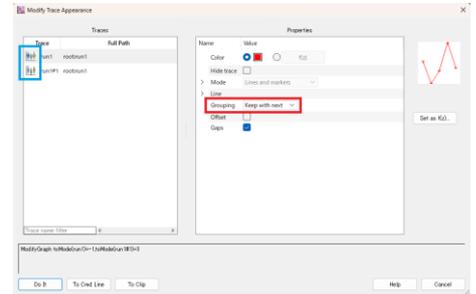
5. メニュー Graph → Append To Graph → Box Plot を選択して、ボックスプロットを追加します (同じ run1 を選択します)。

カテゴリの X 軸を使っているため、プロットは横に並びます。



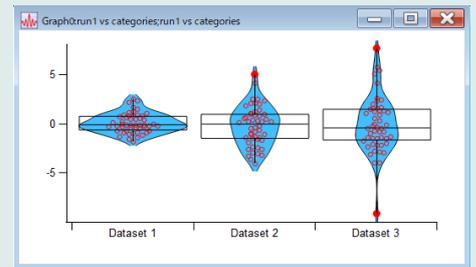
6. バイオリンプロットのトレースをダブルクリックして、Grouping のモードを Keep with next にします。

左のプロットリストには小さなアイコンがあり、トレースがバイオリンプロットなのか、ボックスプロットなのかがわかります。

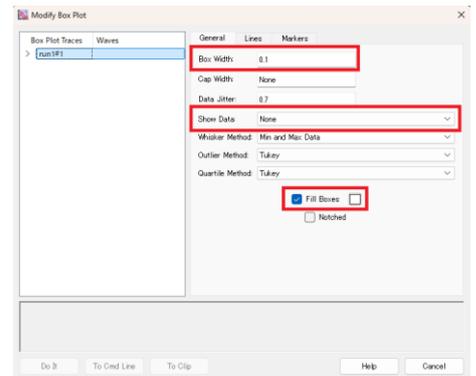


7. 2つのトレースが重なります。最も見栄えが良くなるようにするには、ボックスプロットを狭くし、ボックスを異なる色で塗りつぶすのが良いでしょう。

ここでは、ボックスの幅を 0.1 に設定し、ボックスを白で塗りつぶします。



8. メニュー Graph → Modify Box Plot を選択します。General タブで、Box Width を 0.1 に、Show Data を None に、Fill Boxes チェックボックスにチェックを入れ、色を白にします。



9. データポイントは表示していませんが、必要であれば、塗りつぶしがデータポイントを隠してしまわないように、ボックスプロットのトレースの設定で、表示します。

