

## 3:ヒストグラムと度数分布

このセクションでは、ヒストグラム(度数分布を表現する棒グラフ)の作成方法について解説します。ヒストグラムは次のどちらのデータからでも作成できます。

- 収集したデータからすでに度数を集計したもの
- 集計前の生データ

---

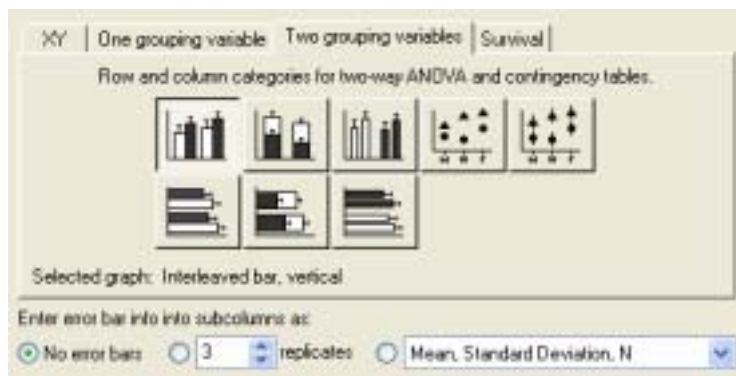
### 集計済みのデータでヒストグラムを作成する

このセクションでは集計済みデータをデータシートに入力し、ヒストグラムを作成する方法について解説します。この場合、ヒストグラムは棒グラフおよびXYグラフのどちらを使っても作成できます。

#### 棒グラフによるヒストグラム

ここでは55人の男性の身長(インチ単位)を測定し、2インチごとにグループ分けしたデータを処理するものとします。

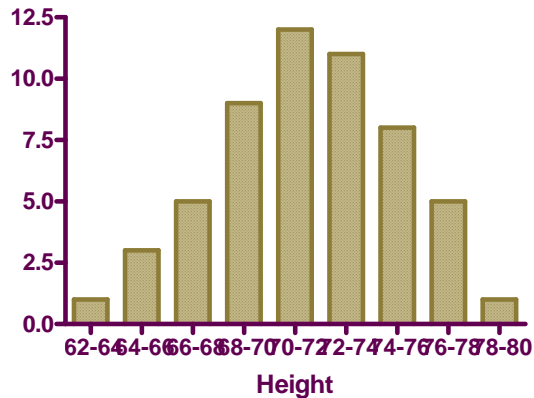
Welcome ダイアログで **Type of graph** を選択し、図に示すように **Two grouping variables** タブで左上のグラフアイコンをクリックします。



OK ボタンをクリックしてデータシートが表示されたら次のページのようにデータを入力します。

	X Labels	A
	Height	Frequency
	X	Y
1	62-64	1
2	64-66	3
3	66-68	5
4	68-70	9
5	70-72	12
6	72-74	11
7	74-76	8
8	76-78	5
9	78-80	1

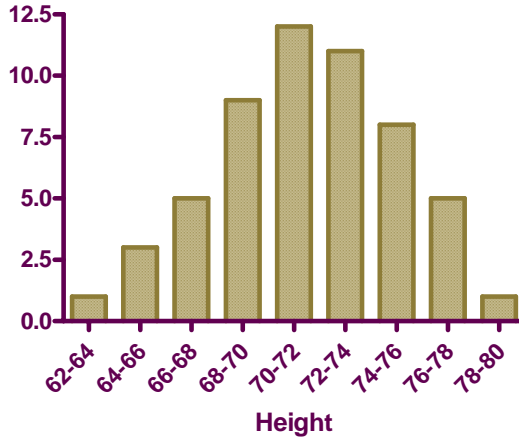
ツールバーにある黄色の **Graphs** タブをクリックすると、次のヒストグラムが画面に表示されます。



作成された図では X 軸のラベルが重なっています。よって、X 軸をダブルクリックして **Format Axes** ダイアログの **X axis** タブを表示し、**Numbering or labeling** の項目を次のように変更します。



ラベルを回転したら、X 軸タイトルの「Height」の周辺にマウスを移動し、両端に矢印が表示されたらタイトルを下方方向にドラッグします。ラベルやタイトルが読みやすくなります。



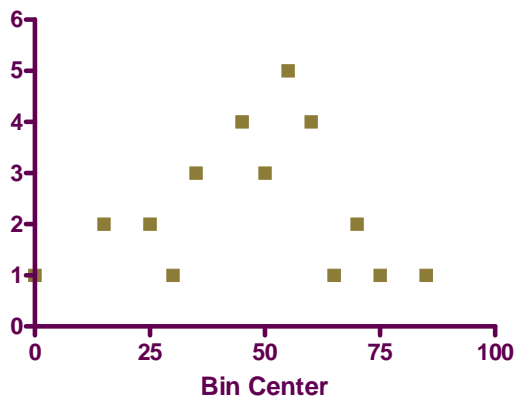
既に集計した値が手元があれば、このように比較的簡単にヒストグラムを作成できます。棒ラベルは文字として予め入力されていたものですから、見栄えを整える作業も楽に行えます。棒ラベルを水平な位置に保っておきたいような時はベースラインをドラッグして、グラフ全体の幅を広げます。または、**Change... Selected Text...**としてフォントサイズを小さなものに変更します。また、ラベル付けの間隔を変更することによって体裁を整える方法もあります。ラベル付け間隔の変更は、この例のようなラベルよりも、単純な数値ラベルの場合に用いる方が効果的です。特定のXラベルを削除したいような場合は、X列から目的のラベルを削除し、Y値だけを残します。

### XY グラフによるヒストグラム

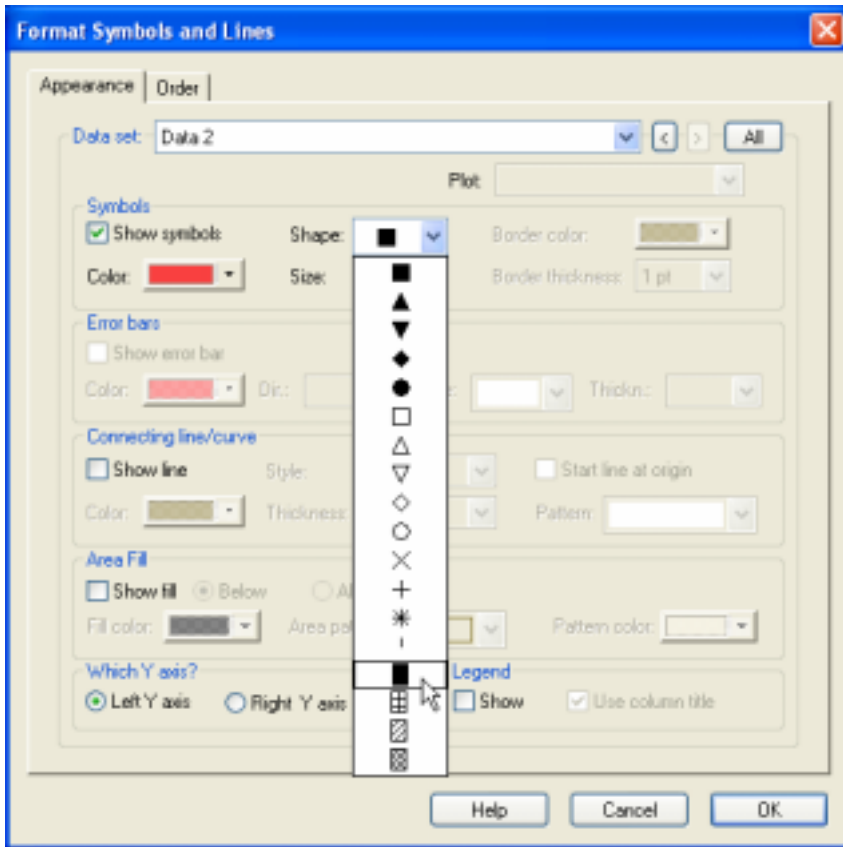
集計済みデータからヒストグラムを作成する方法を、もうひとつ紹介します。先ほどの例に比べると操作量はやや増えますが、ラベル付けの自由度が向上します。つまり、離散的な棒とラベルを中央で位置揃えできますし、線グラフをスーパーインポーズすることもできます。黄色の **Data** タブをクリックし、さらに **New** ボタンをクリックします。そしてプルダウンメニューから **New Data Table (+Graph)** を選択します。**Create New Table** ダイアログで **Create new table (choose X and Y format)** を選びます。そしてX列の属性を **Numbers (XY Graph)**、Y列の属性を **A single column of values** とします。データシートが表示されたら、次に示すデータを入力します。

	X Values	A
	Bin Center	Frequency
	X	Y
1	0	1
2	15	2
3	25	2
4	30	1
5	35	3
6	45	4
7	50	3
8	55	5
9	60	4
10	65	1
11	70	2
12	75	1
13	85	1

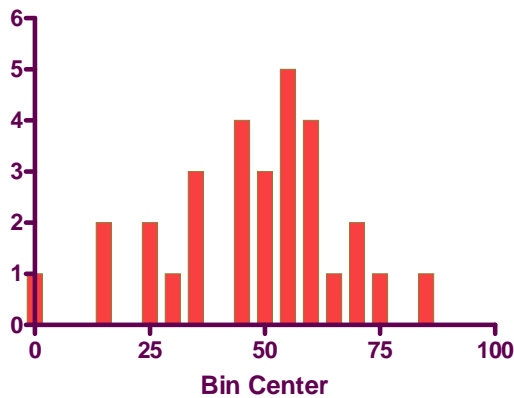
ツールバーにある黄色の **Graphs** タブをクリックします。



グラフ上の記号を点から「spikes」に変更することによって、自動作成された XY グラフを棒グラフに変更します。このグラフはあくまでも XY プロットですから、ダイアログに表示される軸に関する設定は、棒グラフのそれとは異なります。グラフ画面で **Change.. Symbols & Lines** と操作します。そして **Appearance** タブで記号を下から 4 番目にある「spikes」に変更します。



棒の太さを変更する場合は記号のサイズを変更します。



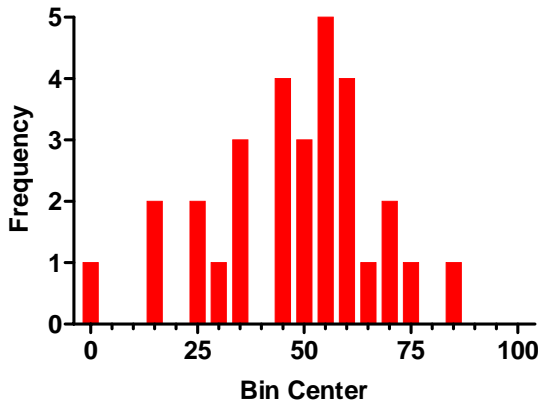
この方法で棒グラフを作成するとラベルと棒が中央揃えで表示されるので、とても見やすいグラフになります。また、棒の間隔も適度に確保されています。しかし、左端の棒が軸と重なっていますので、これを変更します。ベースラインをダブルクリックします。X axis タブの Range と Tick options の項目を次の図のように変更します。Range の Auto チェックボタンを外します。このようにして、Y 軸の右側にオフセット領域を確保します。また、Minor ticks のドロップダウンリストで 5 を選択して副軸目盛を付け加えます。



ついでに Y 軸の表示範囲を変更します。同じ Format Axes ダイアログで Y axis タブを選択します。そして Range と Tick options の設定を次のように変更します。



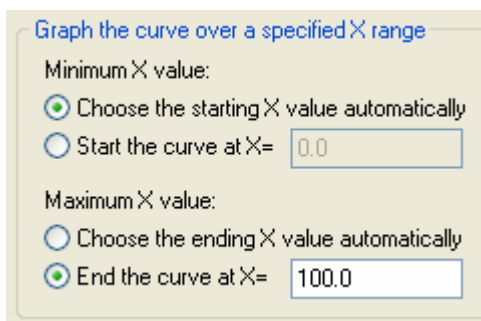
Y 軸のタイトルも図のように変更します。



## ガウス分布の理論曲線を追加する

プロットした棒グラフに正規分布曲線を追加します。画面上のボタンを使って **Analyze... Curves & regression... Nonlinear regression (curve fit)** と操作します。

**Parameters: Nonlinear Regression (Curve Fit)** ダイアログでは of **Classic equations** のリストで **Gaussian distribution** を選択します。デフォルトで正規曲線はデータ範囲内だけにプロットされますので、ここでは X 軸の範囲全体に曲線をプロットするよう範囲を変更します。 **Range** タブで次のように設定します。



Graph the curve over a specified X range

Minimum X value:

Choose the starting X value automatically

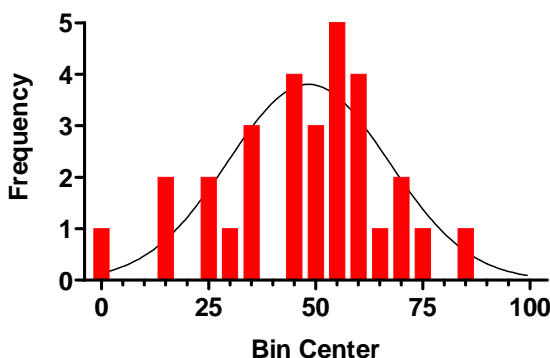
Start the curve at X=

Maximum X value:

Choose the ending X value automatically

End the curve at X=

OK ボタンをクリックすると、次のような曲線がプロットされます。

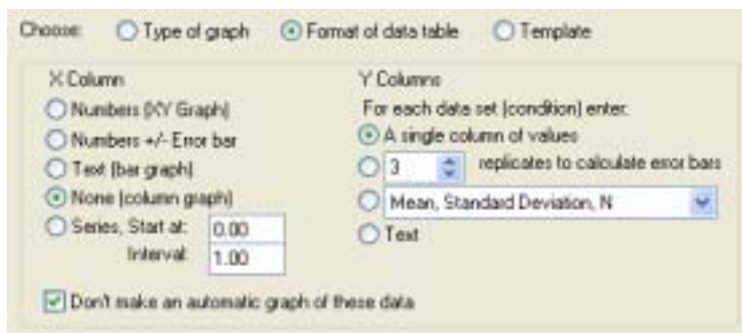


---

## 自動集計機能を使ってヒストグラムを作成する

集計前のデータから自動的に度数分布を求め、ヒストグラムを作成する機能について解説します。

Welcome ダイアログで次の図のように **Format of data table** を選択します。



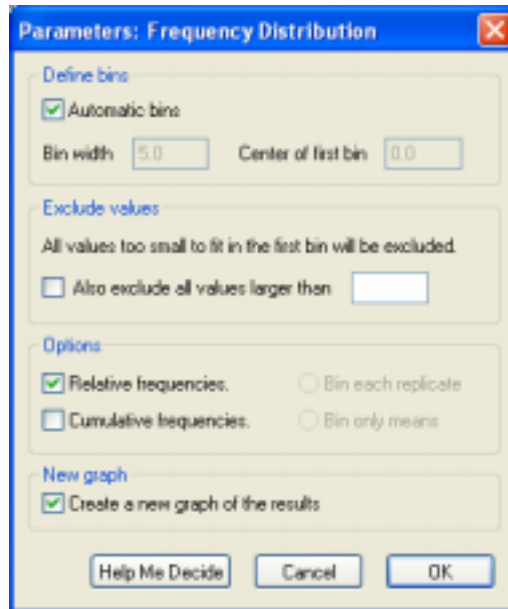
データテーブルで、ひとつの Y 列につぎのようにデータを入力します。

	A
	Data Set-A
	Y
1	30
2	17
3	22
4	19
5	23
6	17
7	22
8	19

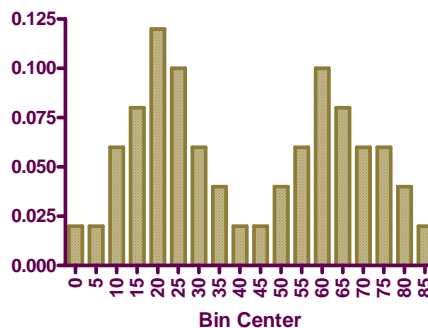
ここで利用するサンプルデータは次の通りです。全部で 50 個のデータがあります。

30 17 22 19 23 17 22 8 19 49  
 37 28 61 18 60 46 74 69 4 61  
 23 71 66 24 42 78 64 60 70 63  
 83 74 2 31 57 20 80 23 15 57  
 16 24 9 58 67 10 54 35 52 76

Analyze ボタンをクリックし、**Statistical analyses** のリストから **Frequency distribution** を選択します。次に示す **Parameters: Frequency Distribution** ダイアログボックスが表示されます。



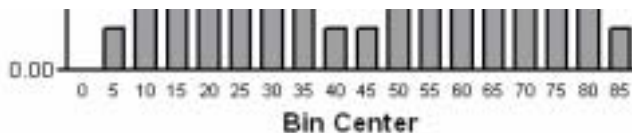
**Define bins** の項目では階級の作成方法を設定します。Prism による自動設定とユーザによる階級幅の設定が可能です。**Exclude values** の項目では外れ値の設定を行います。Prism は一番小さな階級よりも、さらに小さい値はデータ分析の対象から除外します。同様に、任意の値よりも大きなデータを外れ値として除外する場合は、オプションをチェックし、値を入力します。**Options** では階級に含まれるデータ数の絶対値をプロットするか、または全サンプルに対する割合をプロットするかを設定します。**Relative frequencies** は割合をプロットする場合に選択します。また、それらの累積値をプロットする場合は **Cumulative frequencies** を選択します。上図のようにオプションを設定すると、(かならず **Create a new graph of the results** を選択してください)次のようなグラフが作成されます。黄色の **Graphs** タブをクリックして確認しましょう。



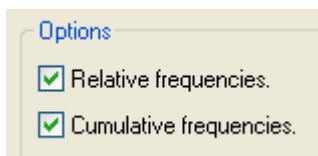
既に解説したようにベースラインをドラッグすると、ラベルの間隔が広がります。



逆にラベルを小さくする場合はベースラインを選択して、**Change... Selected Text...**と操作して文字サイズを変更します。



最後にヒストグラムを、累積値を示すタイプに変更します。グラフを表示した状態で**Change... Analysis Parameters...**とし、オプションを次のように変更します。



追加的な編集を施したグラフを次に示します。

