

# エクセルできれいなヒストグラムを作るには

まとめ: 柳瀬陽介 (2009/12/08)

ここでは**エクセルを使ってきれいなヒストグラム (度数分布表) を作る方法**を解説します。

本当はこういったことは、**ググったり何とかして、自力でやって欲しい**のですが、初心者は小さなつまずきからとてもイライラしてしまい「エクセルやコンピュータなんて嫌い!」となってしまうがちなので、こちらで比較的丁寧に解説します。

**しかし私はこのような解説を示すことで、皆さんの学ぶ力を深いところで損ねているのではないかという懸念を払拭することができません。**

どうか皆さん、これをきっかけに自力でエクセルやコンピュータについて学習する力をつけて下さい。

このヒストグラム作成法で学ぶべきことは、エクセルのバージョン変更ごとによって変わってしまうような具体的手順ではなく、エクセルに慣れて自学自習できるようになることです。**将来ヒストグラムを作成するとかしないとかで考えるのではなく、この課題を通じて自ら学ぶ力をつけて下さい。**

社会に出たら皆さんは他人に親切な人間になってください。しかし現実世界では皆さんが常に他人から親切を期待することはできません。「社会人は自助努力をしろ」が原則です。

学校で、ひな鳥が親鳥から口移しに餌をもらうような学び方ばかりしていた人は社会に出たら自助能力・自学自習能力がないのでとても苦労します。どうぞ「教えてくれないとわかりません」と言った甘えた態度ではなく、今のうちから**「考え、調べ、尋ねる」**態度を身につけておいてください。

[http://yanaseyosuke.blogspot.com/2009/04/blog-post\\_13.html](http://yanaseyosuke.blogspot.com/2009/04/blog-post_13.html)

\*\*\*\*\*

■まず「**1 組第 1 回テスト**」のデータを使ってヒストグラムを作りましょう。データは<http://www.box.net/shared/static/7hoqxfka7t.xls> からダウンロードして下さい。

■エクセルの「**データ分析**」を起動します。(Excel 2003 なら「ツール」から、2007 なら「データ」から「データ分析」を選択してください。皆さんのパソコンの中には「データ分析」がインストールされていない場合があります。その際はインストール DVD を挿入して「データ分析」を「アドイン」してください)

■テスト得点データ領域以外の任意のセルをクリックして、「データ分析」から「**ヒストグラム**」を選択します。

■一番上の空欄「**入力範囲**」に分析したいデータの範囲を指定します。範囲はマウスをドラッグして指定するやり方と、空欄に (絶対セル参照形式で) セルの番地を指定するやり方があります (例 \$B\$2:\$B\$31)。

■「出力オプション」は、一番上の「**出力先**」をチェックして、任意のセル (データ領域以外) を指定しておいてください。こうするとデータがあるシートと同じシート (同じ画面) にヒストグラムが出てきて何かと便利です。

■下の「**グラフ作成**」を**必ず選択**してください。ここをチェックしておかないとヒストグラムは出てきません。

■その他の欄は空欄にしておいてください。実は上から二つ目の空欄である「データ区間」は重要なのですが、後で説明します。

■OK を選択してください。「出力先」で指定したセルから、「**データ区間・頻度**」の表と「**ヒストグラム**」の図が現れるはずです。

■しかしこのヒストグラムの「**データ区間**」は、例えば「13, 28.6, 44.2, 59.8, 75.4, 次の級」になるなど、**わかりにくい**ものになっています。「28.6 のデータ区間の頻度が 2 である」と言われても何のことかわからないのではないのでしょうか。

■これはエクセルが自動的・機械的にデータ区間を計算して決定したから生じています。私たちの感覚からは、データ区間は「0 点台、10 点台、20 点台・・・90 点台」である方がわかりやすいはずです。

■ではどうすればヒストグラムをそのように変えることができるのでしょうか。

■簡単な方法は、**自分で「データ区間」をエクセルに入力**することです。「10, 20, 30, 40, 50, 60, 70, 80, 90, 100」の数字を縦に (列方向に) 10 個のセルに入力して下さい

い。(この時、10 と 20 の二つのデータを入力したら、その二つのセルをアクティブにしてその右端を選択してください。+のマークが出たら下にドラッグするとデータが二つのデータの差ごとに増える等差数列の形で**連続入力**されます)

■先ほど出現した「データ区間・頻度」の表と「ヒストグラム」の図を行ごと削除してもう一度ヒストグラムをつくりましょう。(先ほど作った 10 から 100 のデータ区間は残して置いて下さい)

■今度は「入力範囲」「出力先」「グラフ作成」だけでなく「**データ区間**」も**入力**します。空欄を一度クリックしてアクティブにしてから、マウスのドラッグで範囲を指定して下さい。

■今度はデータ区間が 10 単位で増える、少しはまともなヒストグラムができましたね。しかしこのヒストグラムにも問題があります。例えば「データ区間 20」の頻度は「1」となっています。しかし 20 点台のデータは 24, 29, 28 と 3 つあります。

■元のデータとデータ区間・頻度を比べてみましょう。私たちの感覚からすれば 10 点台のデータは「13」の 1 つだけ、20 点台のデータは「24, 29, 28」の 3 つ、30 点台のデータは「30, 32」の 2 つ、40 点台のデータは「44, 43, 48, 45, 48, 41」の 6 です。ところが頻度はそれぞれ 0, 1, 4, 1, 7 となっています。**明らかに私たちの常識と異なる表記となっています!**

■これは「データ区間」で入力した「10」や「20」といった数字の意味合いが、「10 点台」や「20 点台」ではないことから生じているものです。

■**エクセルのヒストグラムの「データ区間」で入力する値は「上限境界値」です。**その数字を最大としたデータの数を示しているのです。

■ですから「**データ区間 10**」の頻度は「 **$x \leq 10$** 」の範囲にあるデータの数を示しています。それに続く「データ区間 20」の頻度は「 **$10 < x \leq 20$** 」の範囲にあるデータの数を示すわけです。(これが「 $x \leq 20$ 」でなく自動的に「 $10 < x \leq 20$ 」になっているところがエクセルの賢いところですね)

■ですから「データ区間 10」のデータは「 $x \leq 10$ 」ですから 0 個です。「データ区間 20」のデータは「 $10 < x \leq 20$ 」ですから「13」の 1 個です。「データ区間 30」は「 $20 < x \leq 30$ 」ですから「24, 29, 28」だけでなく「30」も入るので 4 個です。「データ区間 40」は「 $30 < x \leq 40$ 」ですから「30」は入らず「32」の 1 個だけです。「データ区間 50」は「 $40 < x \leq 50$ 」で「44, 43, 48, 45, 48, 41」に加えて「50」も入り「7 個」となります。

■こうしてみますと**エクセルのヒストグラムの「データ区間」の考え方 (=上限境界値) は、私たちの直観的理解 (〇点台) と異なる**と考えるべきでしょう。注意が必要です。

■では、どのようにして私たちの直観的理解にかなったヒストグラムを作りましょう？

■**エクセルの「上限境界値」の概念を利用しましょう。**テストの点が小数点を含まない整数値だとしたら「0 点台」とは「 $x \leq 9$ 」のことですから「データ区間」の値は「9」にすればいいわけです。同じように「10 点台」とは「 $9 < x \leq 19$ 」ですから「データ区間」の値は「19」にすればいいわけです (データは整数値ですから 9 より大きい最小値は 10 です)。

■**このように「データ区間」は「9, 19, 29, 39 . . . 89, 100」と指定してやれば、ヒストグラムは私たちの考えるところの「10 点台、20 点台 . . . 90~100 点台」の頻度を示してくれます。**(「データ区間」の最後の値は 99 でなく 100 であることの意味はおわかりですよ。99 にしてしまうと 100 点満点のデータが入らなくなってしまいます)

■データが整数値だけではなく、小数值も含んでいたらどうしましょう？ しかしテストのデータなどはたいてい小数 1 桁ですよ。そうすれば「データ区間」を「9.9, 19.9, 29.9 . . . 」とすればよいだけです。

■それでは「データ区間」を「9, 19, 29, 39 . . . 89, 100」にしてもう一度ヒストグラムを作ってみましょう。今度は私たちの常識的概念にそった頻度が示されます。

■しかしヒストグラムの「データ区間」が「9, 19, 29, 39 . . . 89, 100」さらに「**次の級**」となっているのはいかにも**不格好**ですね。これを修正しましょう。

■修正の仕方は簡単です。「ヒストグラム」の左にある「**データ区間・頻度**」の表の「**データ区間**」の**データを変えます**。「9」を「0 点台」に代えてみて下さい。どうになりました？

■もちろん「0 点台」ではなく「0~9」あるいは「0-10」とやってもかまいません。ただしこのようなデータ入力をする**とエクセルは時にそういうデータを日付データなどと勘違いしてとんちんかんな表示をしてしまいます**。

■こういった場合、入力するデータがどのような種類であるかを指定してやる必要があります。「**セルの書式**」→「**表示形式**」の中の「**文字列**」を選択して下さい

い。これであなたが入力するデータはたとえ一見「数字」のように見えても、エクセルはそれを「数値」とも「日付」とも認識せずに「文字列」と認識して処理します。

■ちなみにこの「セルの書式」はよく使う重要な機能です。「配置」の「**折り返して全体を表示する**」や「**方向**」などは特によく使いますので覚えておいてください。

■さらにヒストグラム(度数分布表)は、通常の棒グラフと違って**棒と棒の間を空けずにつめて表示**することが普通です。ヒストグラムのタイトルなどもこのままでは何のことかわかりません。修正しましょう。

■修正の仕方の詳細を覚える必要はありません。**大原則「困ったときは右クリック」**でいろいろな箇所を右クリックして修正して下さい。

■1組第1回テストのデータ 30 個を使って、**下のようなグラフができるまでエクセルと格闘**ーいやエクセルと遊んでみて下さい。

