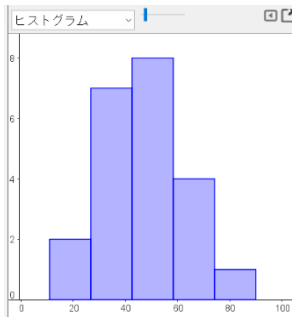
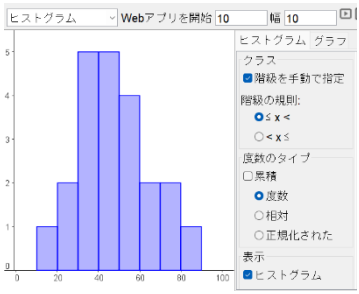
雑感　ヒストグラムと GeoGebra



■　「データの分析」で箱ひげ図や散布図を描くとき，GeoGebra の「表計算・統計」が便利なのでよく利用する．
「1 変数解析」の中にヒストグラムもあるのだが，使い勝手がイマイチのように感じて使うことが少なかった．
　例えば，{11, 25, 28, 31, 33, 37, 36, 38, 41, 44, 47, 45, 49, 50, 53, 54, 55, 65, 66, 70, 71, 90 } という 22 個のデータをヒストグラムにすると，最初，右のような階級数が 5 のものが現れる．[この最初の階級数 5 は，スタージェスの公式※からの $1+\log_2 22 \fallingdotseq 5.45$ によるものかも知れない]
この階級数は，上の青い ▌を動かすことによって変えることができるが，階級の区切りがどうなっていくのかよく分からないので，実用的ではないように思われる．
しかし，◪ をクリックして階級条件を指定すれば，右のように希望の設定が可能になっている（恥ずかしいことに，先日までこれを知らなかった）．

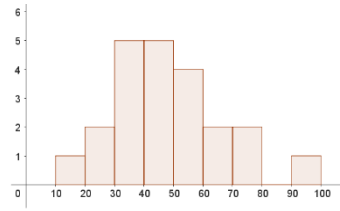


■　ところが，表示されたヒストグラムを子細に見ると，データの内，90 が正しく描かれていない．設定の階級の規則から，本来 90 は階級 $90 \leqq x < 100$ に長方形が描かれるべきデータであるが，階級 $80 \leqq x < 90$ に描かれてしまっている．
　[この 90 を 91 などに変更してみると，なぜか正しく描いてくれる]
このエラーと思われる現象は，Yahoo 知恵袋で質問の形で指摘されたものである．
回答で参照サイト https://wiki.geogebra.org/en/Histogram_Command を紹介したところ，質問者からは「一番右側の階級は「 $a \leq x \leq b$ 」となるよう」だとの返答があった．実際確認してみると

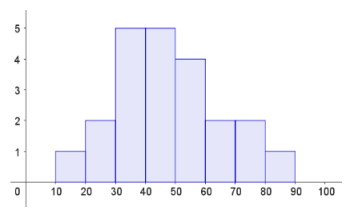
By convention this uses the a ≤ x < b rule for each class except for the last class which is a ≤ x ≤ b

とある．これは，100 点満点の試験のとき，満点の 100 点を 90 点台の階級に含めてしまうような流儀なのであろう．90 でなく 91 だと正しいヒストグラムになるのは，上の注記では不十分な理由が何かあるのだろう．

■　そういった意味では，こういった微妙なルールを知っていないと間違えてしまうことになり，要注意である．
では，GeoGebra ではこの「正しい」ヒストグラムは描けないのかと言うと，面倒を覚悟すればコマンド
Histogram({10, 20, 30, 40, 50, 60, 70, 80, 90},{11, 25, 28, 31, 33, 37, 36, 38, 41, 44, 47, 45, 49, 50, 53, 54, 55, 65, 66, 70, 71, 90}, false)
で可能であり，グラフィック画面に右のように正しく描かれる．



2 つ目の正しくないヒストグラムは，どうやら
Histogram({10, 20, 30, 40, 50, 60, 70, 80, 90},{11, 25, 28, 31, 33, 37, 36, 38, 41, 44, 47, 45, 49, 50, 53, 54, 55, 65, 66, 70, 71, 90}, **true**,10)
といったコマンドで描かれるように仕組まれているようである(右図).



■　奥が深く，知らないことが多い．

※　最近，「スタージェスの経験則」といった記述を見かけるが，彼の名誉のために言っておけば，経験則ではないはず（hist.pdf (plala.or.jp)）．

2023 年 11 月 01 日