1. ヒストグラム

何かしらの問題を解いていると、同じものがいくつあるのか、調べているものの種類は何種類なのかを調べたいときがある。

そんなときは、度数分布を配列で管理すればよい。(この度数分布のことを、慣用的に「ヒストグラム」だとか「バケット」だとか「ビン」と言ったりする)。

配列をhistとしたとき、配列の各要素に対して次のような意味付けをする。

hist[i] = (iに対応する種類のものがいくつあるか)

手始めにこんな問題を考えよう。

N人の身長が与えられるので、そのヒストグラムを出力するプログラムを作る。

入力値はNと、N人の身長(cm)である。これらはすべて整数値である。

ヒストグラムは横軸を度数、縦軸を階級とする。度数はアスタリスクの数で表すこと。また、階級の範囲を100～199(cm)、階級幅を10(cm)とする。

例:

|  |
| --- |
| 10 (入力)  150 160 180 153 190 173 162 189 143 110 (入力)  100-109:  110-119:  120-129:  130-139:  140-149: \*  150-159: \*\*  160-169: \*\*\*  170-179: \*  180-189: \*\*  190-199: \* |

最後のヒストグラム出力は慣れていないと難しいかもしれないが、ここはあまり重要ではない。重要なのは度数をどのように表現するか、である。配列histをつくって、次のように考えればよい。

hist[0] = (100~109cmの身長を持つ人の数)

hist[1] = (110~119cmの身長を持つ人の数)

hist[2] = (120~129cmの身長を持つ人の数)

hist[3] = (130~139cmの身長を持つ人の数)

hist[4] = (140~149cmの身長を持つ人の数)

hist[5] = (150~159cmの身長を持つ人の数)

hist[6] = (160~169cmの身長を持つ人の数)

hist[7] = (170~179cmの身長を持つ人の数)

hist[8] = (180~189cmの身長を持つ人の数)

hist[9] = (190~199cmの身長を持つ人の数)

階級と要素数との規則性を見出せば、次のように一般化できる。

hist[i] = ((100+10\*i)～(109+10\*i)の身長をもつ人の数);

また、うまく切り捨てを考えると、身長をHとしたとき、その身長に対応する配列の添え字は

(H – 100) / 10

で表せる。

このことを踏まえて、目的のプログラムを書くと以下のようになる。

|  |
| --- |
| #include <stdio.h>  int main(void) {      int hist[10] = {};      int N;      int i;      scanf("%d", &N);      for (i = 0; i < N; i++) {          int H;          scanf("%d", &H);          hist[(H - 100) / 10]++;      }      for (int i = 0; i < 10; i++) {          printf("%d-%d: ", 100 + 10\*i, 109 + 10\*i);          for (int j = 0; j < hist[i]; j++) {              printf("\*");          }          printf("\n");      }  return 0;  } |

繰り返すが、このように度数分布を作って問題を解く場合が割とよくある。このとき、先の例のようにうまく配列の添え字とデータをうまく繋ぎ合わせる必要がある。また、階級幅が1の時もよくある。例えば文字列中のアルファベットを調べたいときは、

hist[0] = (‘a’がいくつあるか)

hist[1] = (‘b’がいくつあるか)

…

hist[25] = (‘z’がいくつあるか)

のように定義すればよい。文字をcとしたとき、その文字に対応する配列の添え字は

c – ‘a’

と計算できるのだが、このあたりについての詳しい話は別の機会に述べる。