

## Jで電卓並みの統計ツールをーその2 ー生まのデータからグラフ表示、さらに偏差値の活用ー

西川 利男

統計学を教科書に沿って学ぶというより、Jをツールとして電卓並みに使って、手軽に統計学をやってみようということで始めている。[1]

[1] 西川利男「Jで電卓並みの統計ツールをー統計学はなぜ分かり難いのかー」

JAPLA 研究会資料 2017/12/9

先月のJAPLAの例会では、山本洋一氏よりz値（偏差値）と $\chi^2$ 値との数値的関連性指摘の発表があった。

これに触発されて、今回、私自身もあらためて、偏差値というテーマを見直してみた。

### 1. 生データから、度数分布表、ヒストグラムを作る

高度の統計手法を用いる以前に、得られた生まのデータから度数分布表、そしてヒストグラムを作ることがまず必要になる。これは、統計適用の第一段階で、計算というよりデータの整理にすぎないが、これが思ったより面倒である。

Jはこのような計算以前のこまかい処理が、気軽に行える。

Excelを使えば、いろいろな統計処理計算から、きれいな図やグラフが得られるが、大仰なExcelがなくてはお手上げということは、あまりにバカげている。

これに関連して、かなり古いが、高橋秀俊先生が、「閉じたシステム」と「開いたシステム」という語で、いろいろ書かれている。そのなかで、大きすぎる硬いシステムの危険性を警告されている。[2]

[2] 高橋秀俊「数理の散策」日本評論社(1974), p. 110-121

アメリカ人の大部分は、システムのいろいろあるオプションから選択してやることに慣れすぎて、一部の人を除いて自分で創造することを忘れてしまっている。そして、これがグローバルスタンダード、つまり世界中でどこでも行われていると思っている。

これは、いまや日本人にもあてはまりつつある。このような考え方、生き方が文明化、近代化として、世界中どこにでも蔓延してきている。

Jでのプログラミングは、冷静に人間性を取り戻すことになる、と私は思う。

岩井、鈴木の統計学の本[3]、p. 30 に、課題としてつぎのような、ある学年 80 人の IQ 得点のデータがあげられている。この値を例として、以後進めていこう。

[3] 岩井勇児、鈴木真雄「教師のための統計法入門」福村出版(1989).

NB. 度数値一粗得点

NB. 岩井、鈴木、「統計法入門」p. 53 Q. 2-4

DA =: 100 117 92 88 128 101 94 107 107 68

DA =: DA, 116 135 91 103 58 92 105 113 89 94

DA =: DA, 84 97 104 124 94 107 112 90 97 106

DA =: DA, 95 126 109 105 64 119 99 115 95 112

DA =: DA, 102 75 106 81 77 82 110 90 117 109

DA =: DA, 110 133 126 85 87 109 93 100 86 89

DA =: DA, 96 113 82 76 114 111 91 97 107 100

DA =: DA, 90 72 145 132 101 98 102 101 85 114

DB =: |: 8 10\$DA

DB

100 116 84 95 102 110 96 90

117 135 97 126 75 133 113 72

92 91 104 109 106 126 82 145

88 103 124 105 81 85 76 132

128 58 94 64 77 87 114 101

101 92 107 119 82 109 111 98

94 105 112 99 110 93 91 102

107 113 90 115 90 100 97 101

107 89 97 95 117 86 107 85

68 94 106 112 109 89 100 114

まず、階級にわけて、度数分布表をつくる。階級に区切るには、データの最大値と最小値を見つけて、その間をおよそ 10 段階ぐらいに区切れればよい。

J では、最大値と最小値はつぎのようにして、求められる。

max =: ( {. @¥: ) {}

min =: ( {. @/: ) {}

上の定義は、つぎのとおりである。

最大値(max)は大から小の順のアドレス(¥:)を得て、その先頭({.})の値を使う。

同様にして、最小値(min)は小から大の順のアドレス(/:)で行えばよい。

計算すると、つぎのようになる。

```
max DA
145
min DA
58
```

したがって、 $145 - 58 = 87$  で、10段階に区切れば、9の幅になるが、区切りのよさを考えて、階級の幅を10とする。

つまり、 $140-149, 130-139, \dots, 50-59$  とする。  
さらにこれはそれぞれの中点値、 $144.5, 134.5, \dots, 54.5$  としたほうがよい。

ある値が、どの階級の間隔の間に入るかどうかは、階級の上下で引き算をして、正負の値を調べればよい。Jの次のような動詞を作って調べた。

```
Kyukan =: 50 + 10 * i. 10
test =: 3 : 0
:
q =. x.
p =. 10 + q
P =. p (*@>) y.
Q =. q (*@>) y.
(1 = +/"(1) P,. Q) # q
)

Kyukan
50 60 70 80 90 100 110 120 130 140
Kyukan test 117
110
Kyukan test 92
90
```

さらに、度数分布を調べる動詞 tclass を作った。

```
tclass =: 3 : 0
4.5 + , > Kyukan test"(1 0) L:0 < y.
)
```

上の定義の中で、`test"(1 0)`となっているのは、左引数はランク1 (=ベクトル)、右引数はランク0 (=スカラー)、つまり、それぞれの個々の値に対して、実行するためのものである。

```
Kyukan test 117, 92
110 90
```

つぎのようにして、調べる。

まず、中点値は

```
Chuuten =: |. 4.5 + Kyukan
```

```
Chuuten
```

```
144.5 134.5 124.5 114.5 104.5 94.5 84.5 74.5 64.5 54.5
```

最初からいくつかの値に対して tclass によりどこに属すかを表形式で調べる。

```
Chuuten =/ tclass 100 117 92 88 128 101 94 107 107 68
```

```
0 0 0 0 0 0 0 0 0 0
```

```
0 0 0 0 0 0 0 0 0 0
```

```
0 0 0 0 1 0 0 0 0 0
```

```
0 1 0 0 0 0 0 0 0 0
```

```
1 0 0 0 0 1 0 1 1 0
```

```
0 0 1 0 0 0 1 0 0 0
```

```
0 0 0 1 0 0 0 0 0 0
```

```
0 0 0 0 0 0 0 0 0 0
```

```
0 0 0 0 0 0 0 0 0 1
```

```
0 0 0 0 0 0 0 0 0 0
```

横方向に合計した値はその頻度数を表す。

```
+/"(1) Chuuten =/ tclass 100 117 92 88 128 101 94 107 107 68
```

```
0 0 1 1 4 2 1 0 1 0
```

この値を、中点値と対比して表す。

```
Chuuten ,. +/"(1) Chuuten =/ tclass 100 117 92 88 128 101 94 107 107 68
```

```
144.5 0
```

```
134.5 0
```

```
124.5 1
```

```
114.5 1
```

```
104.5 4
```

```
94.5 2
```

```
84.5 1
```

```
74.5 0
```

```
64.5 1
```

```
54.5 0
```

全部のデータを用いて、最終的な度数頻度表を作る。

```
Freq =: +/"(1) Chuuten =/ tclass DA
```

```
Table_Freq =: Chuuten ,. Freq
```

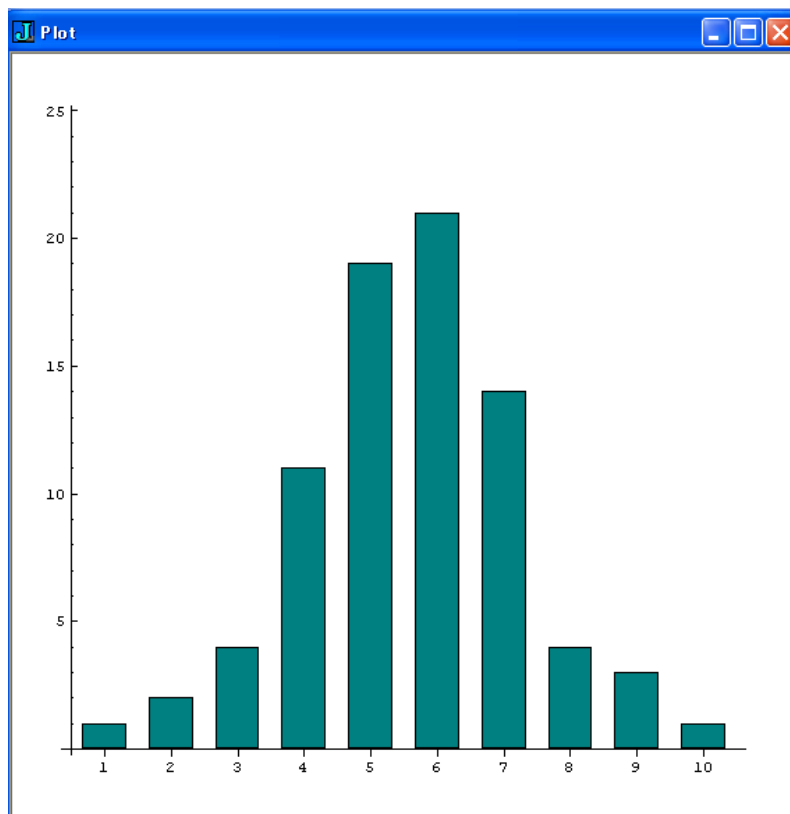
```
Table_Freq
```

```
144.5 1  
134.5 3  
124.5 4  
114.5 14  
104.5 21  
94.5 19  
84.5 11  
74.5 4  
64.5 2  
54.5 1
```

この度数頻度 Freq を用いれば、ヒストグラムは、前回[1]示したように、J の plot ルーチンにより、簡単に次のように表示される。

```
load 'plot'
```

```
'bar' plot |. Freq
```



## 2. 偏差値 = 実効得点、実質得点 を 求める

次に、学校の成績などで使われるいわゆる偏差値について、多少述べたい。これは、平均値、標準偏差などのように、グループの全体のようなすを表す言葉ではない。

偏差値に関連する用語として、次のz値、Z値が定義されている。

$$z = \frac{X - \text{mean}}{\text{std}}$$
$$Z = 10 * \frac{X - \text{mean}}{\text{std}} + 50$$

このZ値が、一般に偏差値といわれているものである。岩井、鈴木 of 統計学の本[3]では、Z得点と名づけられている。

つまり、学校の試験などの各人の得点は、そのままでは、受けた試験の難しさや生徒の出来の良さを反映していない。ところが、このZ得点の値は、平均値、標準偏差を考慮した得点といえる。さらにこのZ得点の値は、100点満点で、標準偏差を50としたときの実質の得点だとしてよい。私は偏差値なる用語の代わりに、各人の実効得点、実質得点という語を使ったら良いと思う。

たとえば実効得点50点ならば受験者の真ん中、70点ならば良い方でその位置もすぐわかる。このような意味から、もっと活用して良いと思う。

z得点とZ得点を求めるJのプログラムは、次のようになる。

```
NB. calc. z and Z scores =====
NB. 岩井、鈴木、「統計法入門」p.53
NB. (mean, stdev) hensha raw score => z(z_score), Z(hensa_chi)
hensa =: 3 : 0
:
'M S' =. x.
z =. (y. - M) % S
Z =. 50 + 10 * ((y. - M) % S)
z,. Z
)
```

平均値、標準偏差値を左引数として、生みの得点それぞれの値（右引数）に対するz得点とZ得点（偏差値）は、次のように求められる。

```
(100.6, 16.44) hensha 144.5 134.5 124.5
2.67032 76.7032
2.06204 70.6204
1.45377 64.5377
```

## プログラム・リスティング

```
mean_std =: 3 : 0
ME =. (+/ > */ L:0 y.) % (+/ > {: L:0 y.)
SD =. %: (+/ > (* / L:0) (*: L:0 ME -~ L:0 ({. L:0 y.)) (,L:0) ({: L:0 y.))
% (+/ > {: L:0 y.)
ME, SD
)
```

NB. stat\_hensachi.ijs

NB. 岩井、鈴木、「統計法入門」 p.30 Q. 2-4,

NB. [7] 偏差値 p. 52-54

```
Kyukan =: 50 + 10 * i. 10
```

```
test =: 3 : 0
:
q =. x.
p =. 10 + q
P =. p (*@>) y.
Q =. q (*@>) y.
(1 = +/"(1) P,. Q) # q
)
```

NB. calc. 度数分布表

NB. 岩井、鈴木、「統計法入門」 p.30 Q.2-4 から 表 2-4

NB. tclass 100 117 92 88 128 101 94 107 107 68

NB. 104.5 114.5 94.5 84.5 124.5 104.5 94.5 104.5 104.5 64.5

```
tclass =: 3 : 0
```

```
4.5 + , > Kyukan test"(1 0) L:0 < y.
```

)

NB. (|. 4.5 + Kyukan) ,. +/"(1) (|. 4.5 + Kyukan) =/ tclass 100 117 92 88

128 101 94 107 107 68

NB. 144.5 0

NB. 134.5 0

NB. 124.5 1

NB. 114.5 1

NB. 104.5 4  
NB. 94.5 2  
NB. 84.5 1  
NB. 74.5 0  
NB. 64.5 1  
NB. 54.5 0

NB. 度数値一粗得点

NB. 岩井、鈴木、「統計法入門」p.53 Q. 2-4

DA =: 100 117 92 88 128 101 94 107 107 68

DA =: DA, 116 135 91 103 58 92 105 113 89 94

DA =: DA, 84 97 104 124 94 107 112 90 97 106

DA =: DA, 95 126 109 105 64 119 99 115 95 112

DA =: DA, 102 75 106 81 77 82 110 90 117 109

DA =: DA, 110 133 126 85 87 109 93 100 86 89

DA =: DA, 96 113 82 76 114 111 91 97 107 100

DA =: DA, 90 72 145 132 101 98 102 101 85 114

DB =: |: 8 10\$DA

NB. (|. 4.5 + Kyukan) ,. +/"(1) (|. 4.5 + Kyukan) =/ tclass DA

NB. 144.5 1

NB. 134.5 3

NB. 124.5 4

NB. 114.5 14

NB. 104.5 21

NB. 94.5 19

NB. 84.5 11

NB. 74.5 4

NB. 64.5 2

NB. 54.5 1

Kyukan =: 50 + 10 \* i. 10

Chuuten =: |. 4.5 + Kyukan

Freq =: +/"(1) Chuuten = / tclass DA

NB. calc. Hensachi



NB. 岩井、鈴木、「統計法入門」 p.53

NB. (mean, stdev) hensa raw score => z(z\_score), Z(hensa\_chi)

NB. (100.6, 16.44) hensa 144.5 => 2.67032 76.7032

NB. (100.6, 16.44) hensa 144.5 134.5 124.5

NB. 2.67032 76.7032

NB. 2.06204 70.6204

NB. 1.45377 64.5377

hensa =: 3 : 0

:

'M S' =. x.

z =. (y. - M) % S

Z =. 50 + 10 \* ((y. - M) % S)

z,. Z

)