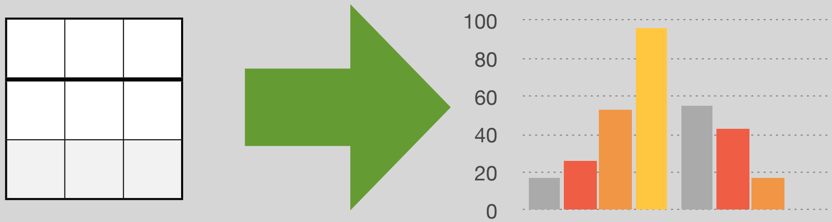
度数分布表の作成方法

author:yuji ookawa

source:fubuki hashimoto

1．度数分布表とは？



各データをある区間ごとに数をかぞえ、そのデータの集まりを表にしたものが”度数分布表”です。ここで言うデータは”数値”です。

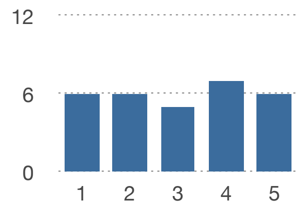
一般的に度数分布表はヒストグラムを作成する下準備として作成されます。ヒストグラムは分布の型を特定したり、データを視覚的に理解するために使用されます。表計算ソフトが使える時代には余計な手間の様に思えます。

しかし、集めたデータを見やすくし、データ解析のための下準備である”データの前処理”を自然と行っている大変重要な作業です。

2．様々な分布の型

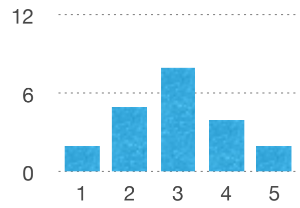
分布の型を特定することは予測をする上で大変重要です。なぜなら、分布の型はそのデータの集まりの性質を表すからです。以下に一例を示します。

均一型：一様分布



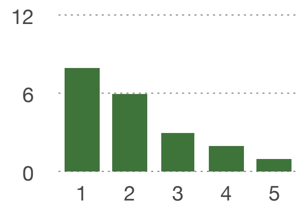
サイコロを振り、出た目をグラフ化するとあらわれる形です。

標準型：正規分布



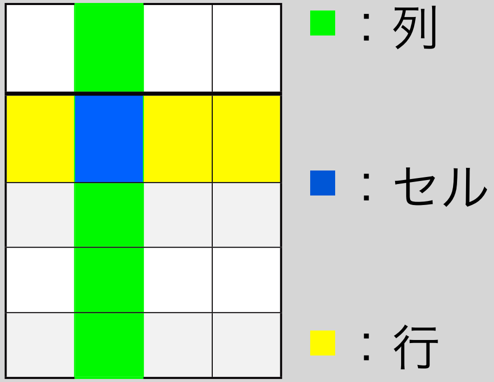
身長や体重などをグラフ化するとあらわれる形です。

斬減型：指数分布



モノの生存数を時間毎にグラフ化するとあらわれる形です。

表の用語解説



統計学では、表のことを”クロス表”と呼び、上の図のような表は”４行x３列のクロス表”と呼びます。（単に４x３のクロス表と呼ぶこともある）

この際、ヘッダー（名目行）とフッター（名目列）は数えません。

セル内の数値は”度数”と呼ばれます。

3-1. 作成方法

最終的に以下のような度数分布表を作成します。赤丸は作成手順を表しています。

元のデータ（n=18）

	1	2	3
A	5	3	9
B	3	6	4
C	2	8	5
D	9	11	7
E	1	2	6
F	8	4	2

8

9

作成する度数分布表

1~7

区間	中心値	度数マーク
3．５～6．５	5.0	
6．５～9．５	8	
9．５～１２．５	11	

計算式の用語解説

平方根 $\sqrt{n}$ ：掛け合わせることで指定された数 n にするための数を求める計算

$n = \sqrt{n} \times \sqrt{n}$

電卓では数値を入力し、

[$\sqrt{}$]を押すことで計算されます。

標本数について

ある母集団からの標本は母集団の性質を完全に引き継いだものとは限りませんが、以下の式で推定の精度を推量することはできます。

$$\frac{1}{\sqrt{n}}$$

3 -2. 作成方法

I.各列に最大と最小の値を見つけ、最大に"○"、最小に"X"を付ける。

	1	2	3
A	5	3	9 ○
B	3	6	4
C	2	8	5
D	9 ○	11 ○	7
E	1 X	2 X	6
F	8	4	2 X

II.Iで見つけた最大、最小のなかで更に最大 L 、最小 S の値を見つけ、同様に印を付ける。

	1	2	3
A	5	3	9 ○
B	3	6	4
C	2	8	5
D	9 ○	11 ◎	7
E	1 XX	2 X	6
F	8	4	2 X

III.区間の数 k を $k = \sqrt{n}$で求める。この際小数点以下は四捨五入する。

$$k = \sqrt{18} = 4.246 \fallingdotseq 4$$

計算式の用語解説

単位 d :値の桁の事

（例）h = 10.56で元の値の桁が3.8のように小数点１であれば、h = 10.6となります。

電卓での分数計算：

$\frac{5-4}{2} = 0.5$の入力手順は

5 - 4 = ÷2

となります。液晶画面では

5 → 4 → 1 → 2 → 0.5

となります。

３-3.　作成方法

Ⅳ.区間の幅 h を $h = \frac{L-S}{k}$で求める。単位 d は元の値に合わせる。

$$h = \frac{11-1}{4} = 2.5 \fallingdotseq 3$$

Ⅴ.第１区間の始点 h^1_{start} を $h^1_{start} = S - \frac{d}{2}$で求める。

$$h^1_{start} = 1 - \frac{1}{2} = 0.5$$

Ⅵ.第１区間の終点 h^1_{end}を $h^1_{end} = h^1_{start} + h$で求める。

$$h^1_{end} = 0.5 + 3 = 3.5$$

Ⅶ.第 n 区間の始点 h^n_{start}を $h^n_{start} = h^{n-1}_{end}$とし、第n区間の終点を $h^n_{end} = h^n_{start} + h$で求め、それを k 回繰り返す。

（例）第２区間の始点は第１区間の終点3.5となり、第２区間の終点は $h^2_{end} = 3.5 + 3 = 6.5$となります。

※区間の数 k が増減しても問題ありません。

３-4. 作成方法

VIII.各区間の中心 h^n_{center}を $h^n_{center} = \frac{h^n_{start} + h^n_{end}}{2}$で求める。VIIの計算結果も追記したものを以下に示します。

区間	始点	終点	中心
1	0.5	3.5	2
2	3.5	6.5	5
3	6.5	9.5	8
4	9.5	12.5	11

IX.上の表を元に元のデータを数えます。度数分布表の度数にはタリースティック（計数表記記号）で記入していきます。最後に度数の合計と元のデータ数が同じであることを確認します。

区間	中心値	度数マーク
０．５〜３．５	2	
３．５〜６．５	5	
６．５〜９．５	8	
９．５〜１２．５	11	

X.完成です。

用語解説

タリースティック：

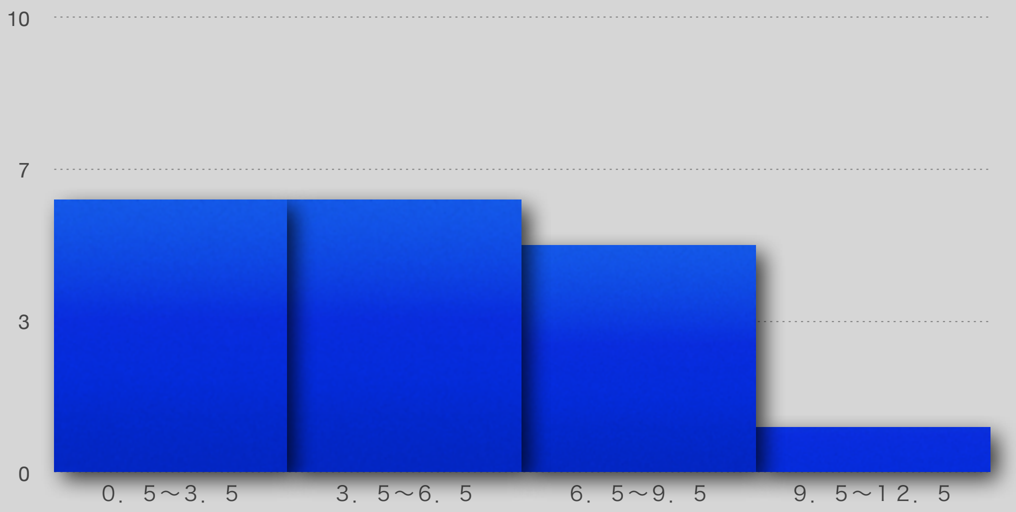
計数する際に、表記する記号のことです。

日本を含む東アジア圏では”正”

西洋圏では が一般的です。

日本は戦後GHQの復興策の一つとしてQCが導入された経緯があるため、QC関連では西洋圏のタリースティックが使われることが多いようです。尚、”正”でも構いません。

追記



作成した度数分布表から簡単なヒストグラムを作成しました。

均一型つまり、一様分布であることがわかります。このデータ群はランダムである（サイコロを投げる事象と同じ）ことがわかりました。ここから言えることは、このデータ群は９．５まではほぼ均一な確率で値が計測されていることがわかり、次回からのヒストグラムに変化があった場合、何らかの変化点があったものと考えることが妥当となります。

因みにこのデータは筆者が思いつきで書いたものでデタラメです。

表紙は工学最大の成果であるスペースシャトルのように飛び立てるようにと選びました。

参考文献＆情報

基本的にはFubuki Hashimoto氏の情報を元にしましたが、補完として以下の文献も参考にしました。

書籍：QC数学のはなし

著者:大村　平

発行：株式会社　日科技連出版社

書籍：統計学入門

著者:盛山　和夫

発行：財団法人　放送大学教育振興会

書籍：LATEX厳選テクニック

著者:竹内　充彦

発行：株式会社九天社