

2016年度秋学期 統計学 第4回

データを「分布」で見る

浅野 晃

関西大学総合情報学部



初学者には、「分布」という考
え方がむずかしいらしい。

鳥居泰彦「はじめての統計学」より

量的データ

ここからは、**量的データ**を扱う

量的データは、**数値の集まり**

データとは、**数値の集まり (集合)**

データに含まれる個々の数値は

「数値」とよぶことにする

分布とは

- データが「分布している」「分布する」
= データが、大小ばらばらの
数値が集まってきている
- 「分布」
= 大小ばらばらの数値が集まって
できたデータ

統計学が相手にするのは「分布」です

度数と度数分布

ばらばらなデータがどんなふう
にばらばらか

どんな値がどのくらい頻繁に
現れるか

「イチロー選手が1試合に打つ
ヒットの数」で言えば、
ヒットの数が0本である試合が
何試合、1本である試合が何
試合、...

度数と度数分布

どんな値がどのくらい頻繁に
現れるか

ヒット の数	[度数] 試合数	[相対度数] 割合
0本	5試合	10%
1本	15試合	30%
2本	10試合	20%
3本	10試合	20%
4本	5試合	10%
計 50試合		100%

[度数
分布]

「測るデータ」の場合は

「身長」は1本、2本と「数える」
のではなく「測る」

数値を「1本、2本」と
ランクにわけることができない

どんな値がどれくらい頻繁に
現れるか、つまり「度数」は
どうやって表すのか？

階級と階級値

測ったデータを、ある間隔で段階に
区切る

	身長	相対度数
	⋮	
	160~165	15%
[階級]	165~170	20%
[階級幅]	170~175	20%
は5cm	175~180	10%
	⋮	

度数分布を作る

階級を決めて、各階級に入る数値を数える

35 62 65 23 40 30 70 55 57 65 15 90 67 65 70 45 80
79 46 45 25 50 62 75 78 48 50 60 75 75 60 78 58 78
63 95 20 46 55 56 70 60 79 18 63 67 85 25 40 50

数えるときには、

/// という記号や 「正」 の字をかく

階級値

以上	未満	階級値	度数	相対度数
15	25	20	4	0.08 (8%)
25	35	30	3	0.06 (6%)
35	45	40	3	0.06 (6%)
45	55	50	8	0.16 (16%)
55	65	60	12	0.24 (24%)
65	75	70	8	0.16 (16%)
75	85	80	9	0.18 (18%)
85	95	90	3	0.06 (6%)
x	x	x	計 50	計 1 (100%)

↑
[階級値]

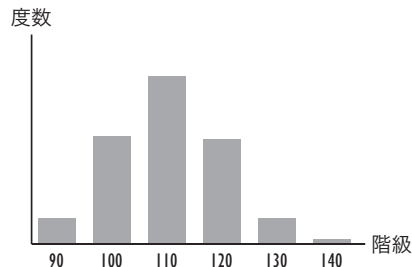
その階級に入った数値は、
どれも概略この値であるとする

ヒストグラム

度数分布をグラフに表す

横軸に階級、縦軸に (相対) 度数

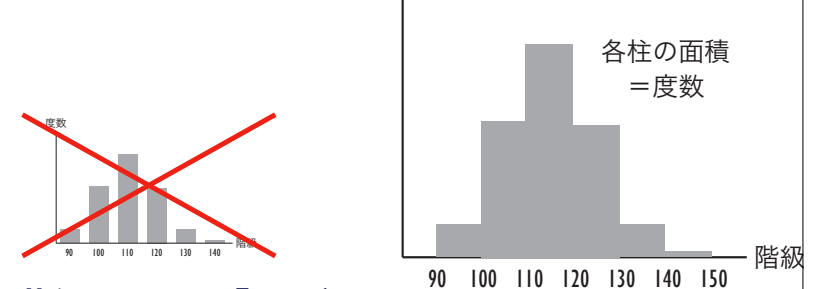
↑
実は微妙に違う



これは
「棒グラフ」

面積で度数を表す

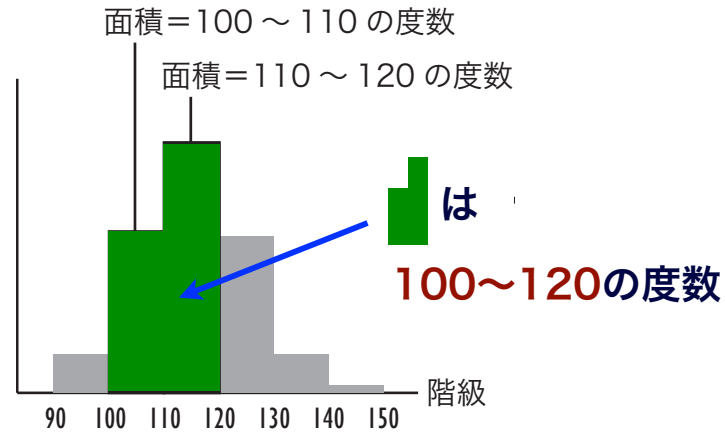
横軸に階級をとり、
階級幅を底辺とする柱の面積で
(相対) 度数を表す



階級の上に「飛び」はない
から、柱の間はあかない

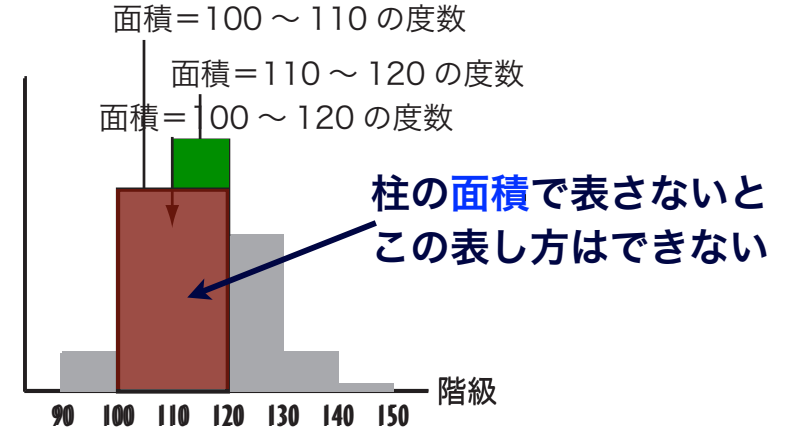
なぜ面積で表すのか

柱をくっつけたり切り分けたりするため



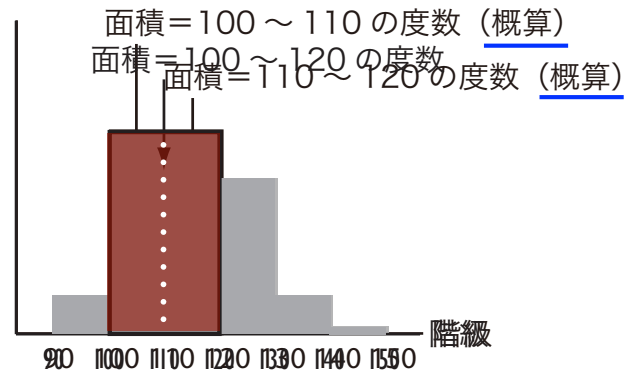
なぜ面積で表すのか

くっつけた柱を1つにしてもよい



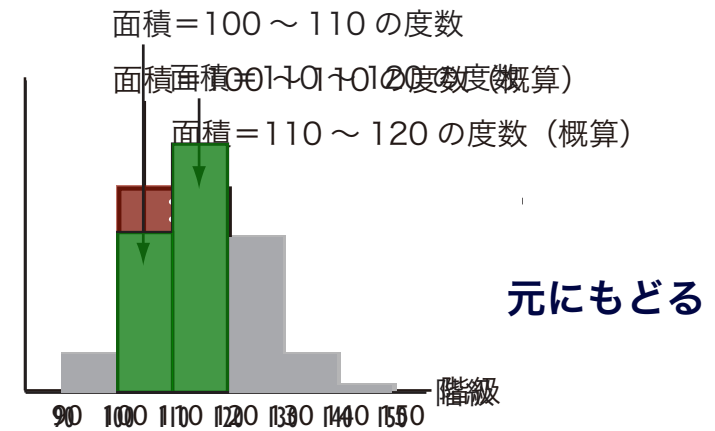
なぜ面積で表すのか

柱を分割してもよい



なぜ面積で表すのか

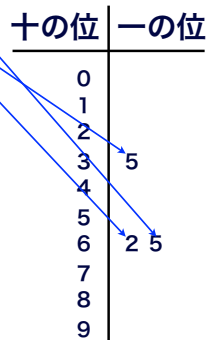
分割した柱を詳細に表すと



幹葉表示

度数が見えるように、数値を並べる

35 62 65 23 40 30 70 55 57 65 15 90 67 65 70 45 80
 79 46 45 25 50 62 75 78 48 50 60 75 75 60 78 58 78
 63 95 20 46 55 56 70 60 79 18 63 67 85 25 40 50

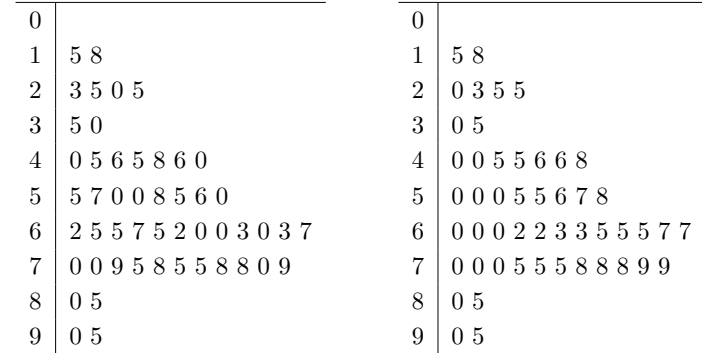


幹葉表示

度数が見えるように、数値を並べる

データ全体を並べると

一の位を昇順に



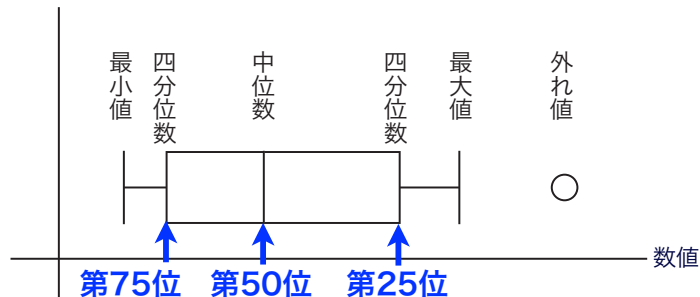
簡易ヒストグラムになり、数値も見える

ボックスプロット

ヒストグラムをさらに簡略にしたもの

順位にもとづく指標

100個の数値からなるデータがあるとして、
 大きさの順に並べ替えた時



パラレルボックスプロット

ボックスプロットは、複数個並べられる

(ヒストグラムにはできない)

