

統計学入門 第2回

早稲田大学政治経済学部
西郷 浩

1

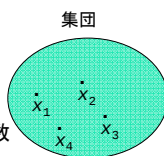
本日の目標

- 度数分布とヒストグラム
 - 度数分布の概念
 - 度数分布表の作成
 - ヒストグラムの作成
 - ヒストグラムの見方
 - 度数分布表における階級の構成
 - (累積度数)分布関数
 - 注: 代表値は次回に説明します。
- PC実習

2

度数分布の概念(1)

- 記号の整理
 - 個体: 集団の構成要素
 - 集団: 個体の集合
 - 変数(特性値): 個体の特徴を示す数量・属性
 - データ: ある集団に属する個体のもつ変数の値を集めたもの
 - $x_i (i=1, 2, \dots, n)$
 - 添え字 i は個体番号
 - n を集団の大きさと呼ぶ。
 - 例: 都道府県別基幹的農業従事者数



3

度数分布の概念(2)

- 変数 x から見た集団の特徴を捉える。
 - 例:
 - 基幹的農業従事者数から都道府県全体の特徴を捉える。
 - どのような方法が考えられるか。

4

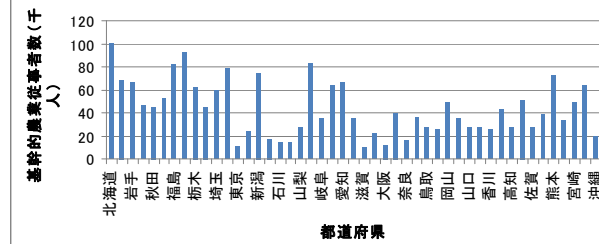
度数分布の概念(3)

101	63	14	10	49	28
69	45	14	23	35	39
67	59	28	12	28	73
46	79	83	40	28	34
45	11	35	16	26	49
52	24	64	36	43	64
82	75	67	28	28	20
92	17	35	26	51	

5

度数分布の概念(4)

図1: 都道府県別基幹的農業従事者数

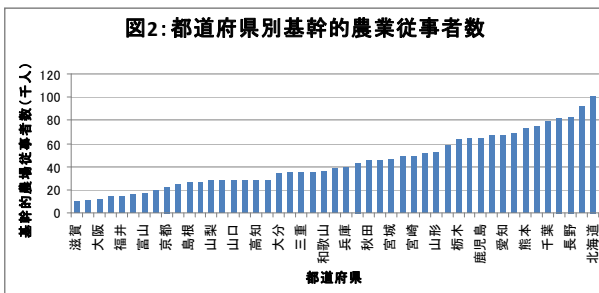


資料: 総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

6

度数分布の概念(5)

図2: 都道府県別基幹的農業従事者数



資料: 総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

7

度数分布の概念(6)

- それぞれの方法の特徴
 - 数字のまま: 特徴がつかみにくい。
 - グラフによる視覚化: 数字のままよりはまし。
 - 並べ替え(sort): 特徴がつかみやすくなる。
 - もっと見やすい方法は? ⇒ 度数分布
 - 変数 x の値をグループ化し、各グループに属する個体数を勘定する。
 - それによって全体的特徴が明らかにできる。

8

度数分布表の作成(1)

表1: 度数分布表

階級 番号	下限 (より大)	上限 (以下)	階級値	度数	相対度数	累積度数	累積相 対度数
1	0	20	10	8	0.170	8	0.170
2	20	40	30	17	0.362	25	0.532
3	40	60	50	9	0.191	34	0.723
4	60	80	70	9	0.191	43	0.915
5	80	100	90	3	0.064	46	0.979
6	100	120	110	1	0.021	47	1.000
合計	--	--	--	47	1.000	--	--

資料: 総務省統計研修所『第61回日本統計年鑑』2012年 表7-3

9

度数分布表の作成(2)

■ 度数分布表の構成

- 階級
 - 変数 x の区間(下限と上限)で指定したグループ
- 階級値
 - $(\text{下限} + \text{上限})/2$
- 度数
 - 階級に属する個体数

10

度数分布表の作成(3)

■ 分布の表し方

- 度数
- 相対度数: $\text{度数}/n$
- 累積度数: その階級までに登場した度数の和
(その階級の上限以下の値をもつ個体数)
- 累積相対度数: $\text{累積度数}/n$
- 目的に応じて使いやすいものを利用する。

11

度数分布表の作成(4)

■ 階級の構成

- 必要条件
 - 排他的(重複ナシ)
 - 網羅的(どこかには入る)
 - つまり「どの個体もどれかひとつの階級に属する」
- 望ましい条件
 - 分布の特徴がつかみやすいように
 - どうやればよいか? (ヒストグラムの見方を終えてから)

12

ヒストグラムの作成(1)

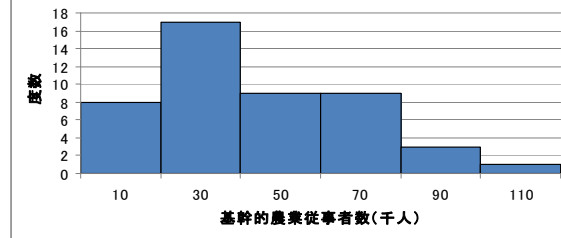
■ ヒストグラム

- 度数分布表を柱状グラフで表したもの。
 - 縦軸: 度数密度(ないし度数)
 - 横軸: 変数 x の値(階級値または下限・上限)
- 原則
 - 柱の面積(底辺 × 高さ)が度数に比例する。
 - 底辺 = 上限 - 下限 = 階級幅
 - 高さ = 度数密度
 - すべての階級について階級幅が同一の場合
 - 柱の高さを度数そのもので置き換えてもよい。

13

ヒストグラムの作成(2)

図3: 基幹的農業従事者数ヒストグラム



資料: 総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

14

ヒストグラムの作成(3)

■ さまざまなヒストグラム

- 縦軸が相対度数(密度)となっているもの。
 - 密度:
 - 相対度数を階級幅で除した値
 - 縦軸を密度にすれば、「柱の面積が度数に比例する」という条件は自動的に満たされる。
- 柱の頂点にあたる部分を直線(ないし曲線)で結んだもの(度数多角形)。

15

ヒストグラムの見方(1)

■ 峰の数

- 単峰分布: 単一
- 多峰分布: 複数(2つの場合: 双峰分布)
 - もととの集団が同質的でない可能性を示唆している。

■ (左右の)対称性

- 対称
- 非対称(歪んでいる):
 - 右に歪んでいる。
 - 左に歪んでいる。

16

ヒストグラムの見方(2)

- 中心の位置
 - 分布の中心はどのあたりか。
- 散らばり具合
 - 分布の幅はどれぐらいか。
- 留意点
 - ヒストグラムの形状(見た目の印象)は階級構成に左右される。

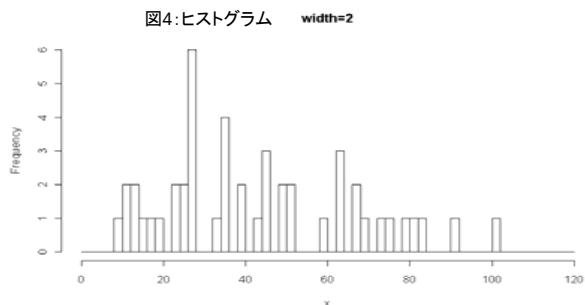
17

度数分布表における階級の構成(1)

- 階級幅の大小が度数分布表におよぼす影響
 - 度数分布表における階級幅: 狭 ←→ 広
 - 度数分布表のふくむ情報量: 大 ←→ 小
 - ヒストグラムの形状の平滑度: 小 ←→ 大
- (初級段階での)現実的な階級幅の決め方
 - 階級構成(階級幅など)を変えて複数のヒストグラムを作成する。
 - 情報量と平滑度の妥協点を見出す。
 - 恣意性の混入は避けられない。

18

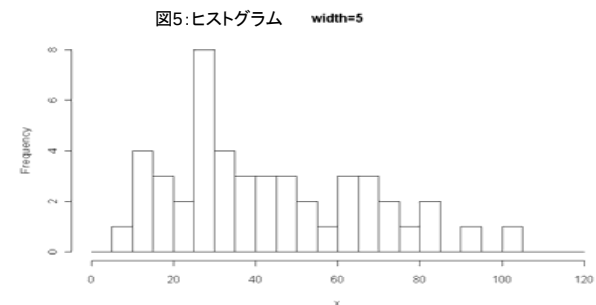
度数分布表における階級の構成(2)



資料:総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

19

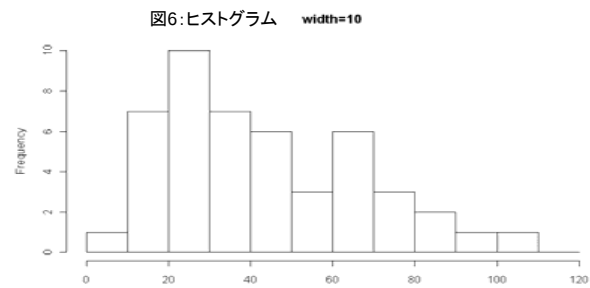
度数分布表における階級の構成(3)



資料:総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

20

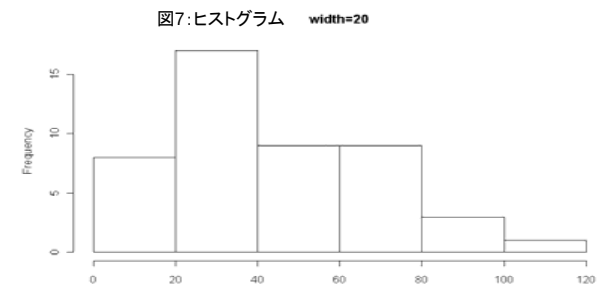
度数分布表における階級の構成(4)



資料: 総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

21

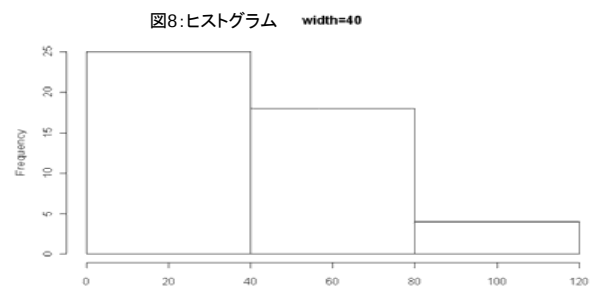
度数分布表における階級の構成(5)



資料: 総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

22

度数分布表における階級の構成(6)



資料: 総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

23

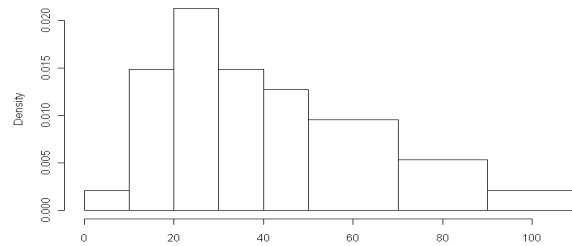
度数分布表における階級の構成(7)

- 階級によって階級幅を変更する。
 - データが密なところ: 狭くする。
 - データが疎なところ: 広くする。
 - 例: 都道府県別基幹的農業従事者数
 - 0千人より大で50千人以下: 階級幅 10千人
 - 50千人より大 : 階級幅 20千人

24

度数分布表における階級の構成(8)

図9:ヒストグラム (異なる階級幅が混在する場合)



資料:総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

25

(累積)分布関数(1)

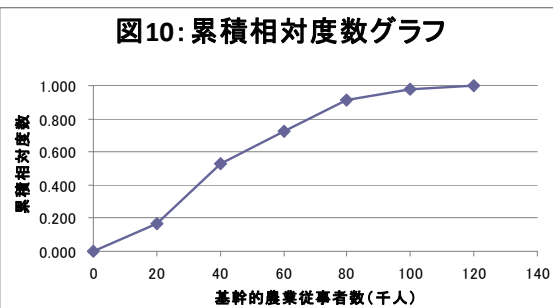
■ 分布関数

- 横軸:階級の上限
- 縦軸:累積相対度数(ないし累積度数)
 - 本来は、「横軸の値よりも小さい値の変数をもつ個体の比率」を表したものが分布関数である。
 - 度数分布表がなくても、元のデータから描くことができる。
- ヒストグラムと分布関数との関係
 - 図2.11(教科書 p. 25)

26

(累積)分布関数(2)

図10: 累積相対度数グラフ



資料:総務省統計研究所『第61回日本統計年鑑』2012年 表7-3

27

(累積)分布関数(3)

■ 分布関数の利用方法

- 横軸→縦軸
 - 横軸の目盛り以下の値の変数をもつ個体の割合(相対累積度数)がわかる。
- 縦軸→横軸
 - 個体を変数の昇順(小さい順)に並べたとき、縦軸の目盛りの割合の個体に対応する変数の値(分位点)がわかる。
 - 中央値や四分位点(後述)の計算に利用できる。

28



PC実習

- 度数分布表の作成
 - frequency(data部分, 上限)
 - Ctrl + Shift + Enter
- ヒストグラムの作成