

統計学I

早稲田大学政治経済学術院

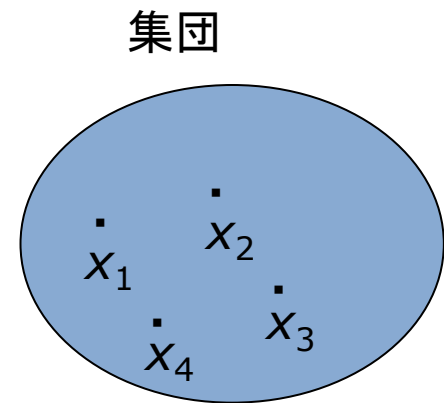
西郷 浩

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
 - 累積相対度数のグラフ(分布関数)
- PC実習

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: 東京都市区町村別世帯数(H27)



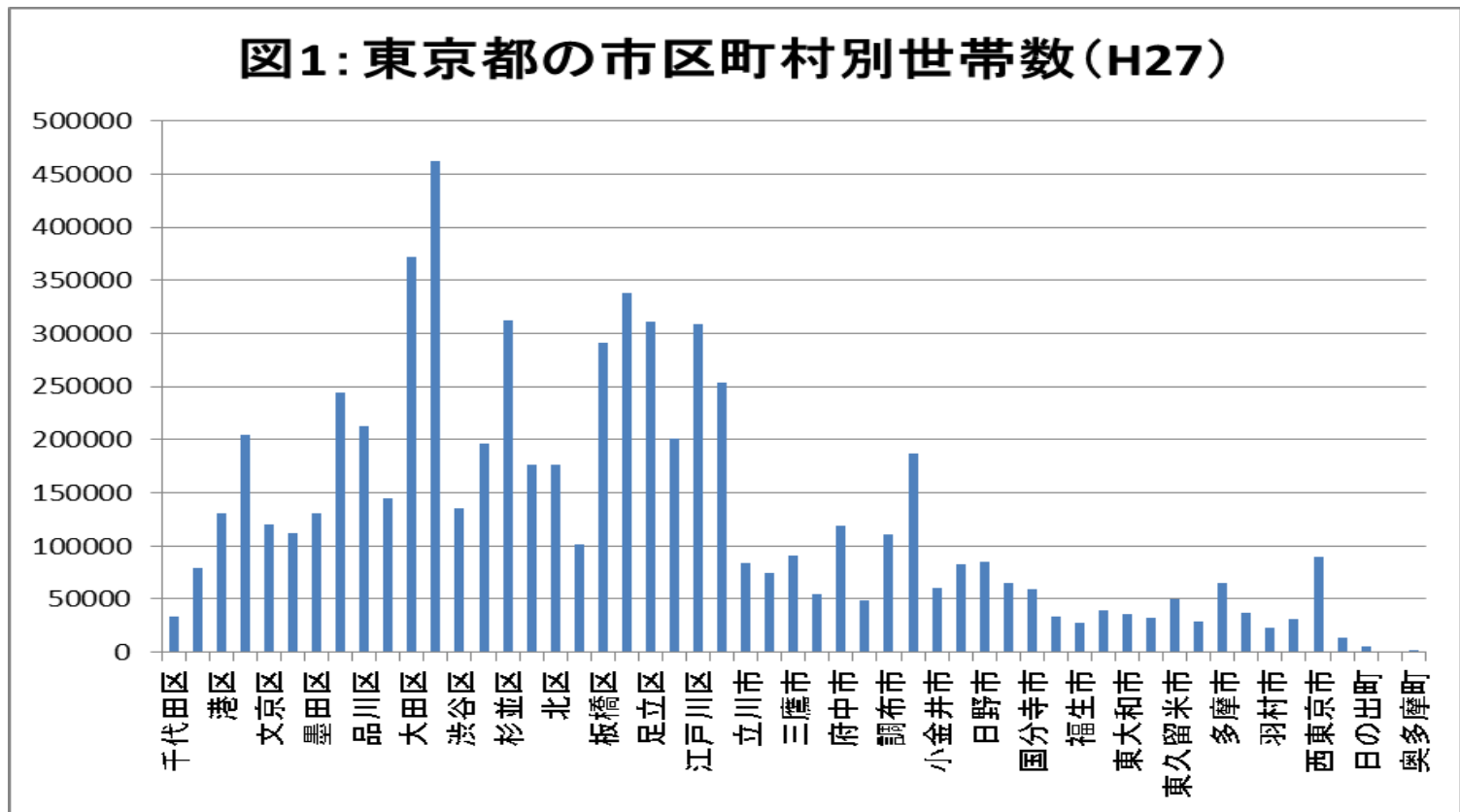
度数分布の概念(2)

- 変数 x から見た集団の特徴を捉える。
 - 例：
 - 市区町村別世帯数から東京都(島嶼部を除く)の特徴を捉える。
 - どのような方法が考えられるか。

度数分布の概念(3)

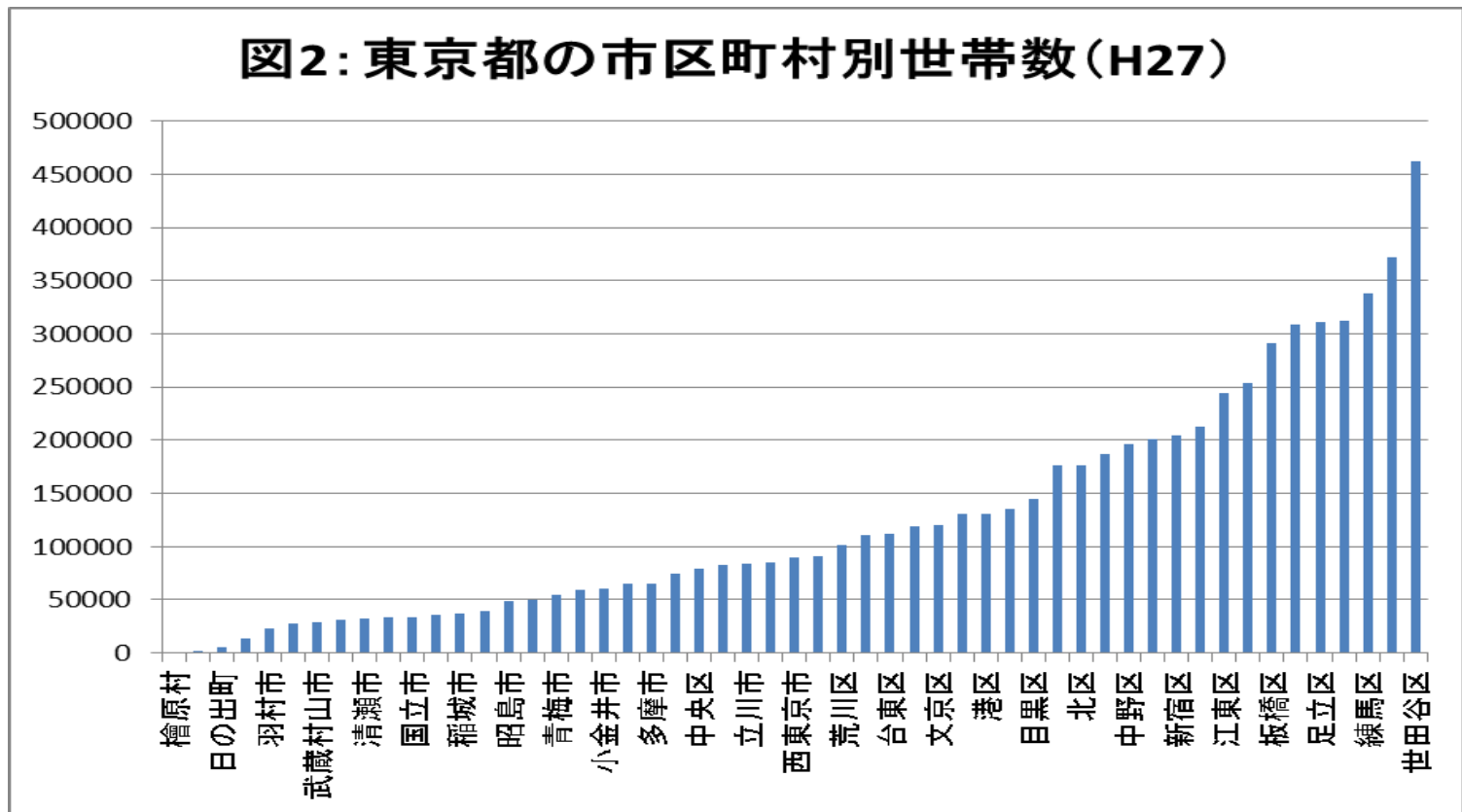
33361	212322	176873	83355	59859	35570	89676
79184	144873	101752	73981	82994	32352	13160
130386	371447	291659	90396	84473	49904	5773
203831	462335	337824	54344	64578	28266	837
120640	135498	310852	119391	59116	65476	2044
111887	195630	201360	48156	33627	36324	
130338	312413	308328	110155	27413	23419	
244482	176104	254008	186376	39520	30734	

度数分布の概念(4)



資料: 総務省「平成27年国勢調査」人口速報集計結果

度数分布の概念(5)



資料: 総務省「平成27年国勢調査」人口速報集計結果

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? $\Rightarrow$ 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。
 - それによって全体の特徴が明らかにできる。

度数分布表の作成(1)

表1: 度数分布表							
階級	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	50000	25000	16	0.302	16	0.302
2	50000	100000	75000	12	0.226	28	0.528
3	100000	150000	125000	9	0.170	37	0.698
4	150000	200000	175000	4	0.075	41	0.774
5	200000	250000	225000	4	0.075	45	0.849
6	250000	300000	275000	2	0.038	47	0.887
7	300000	350000	325000	4	0.075	51	0.962
8	350000	400000	375000	1	0.019	52	0.981
9	400000	450000	425000	0	0.000	52	0.981
10	450000	500000	475000	1	0.019	53	1.000
合計	--	--	--	53	1.000	--	--

資料: 総務省「平成27年国勢調査」人口速報集計結果

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

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。

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)

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 $\times$ 高さ)が度数に比例する。

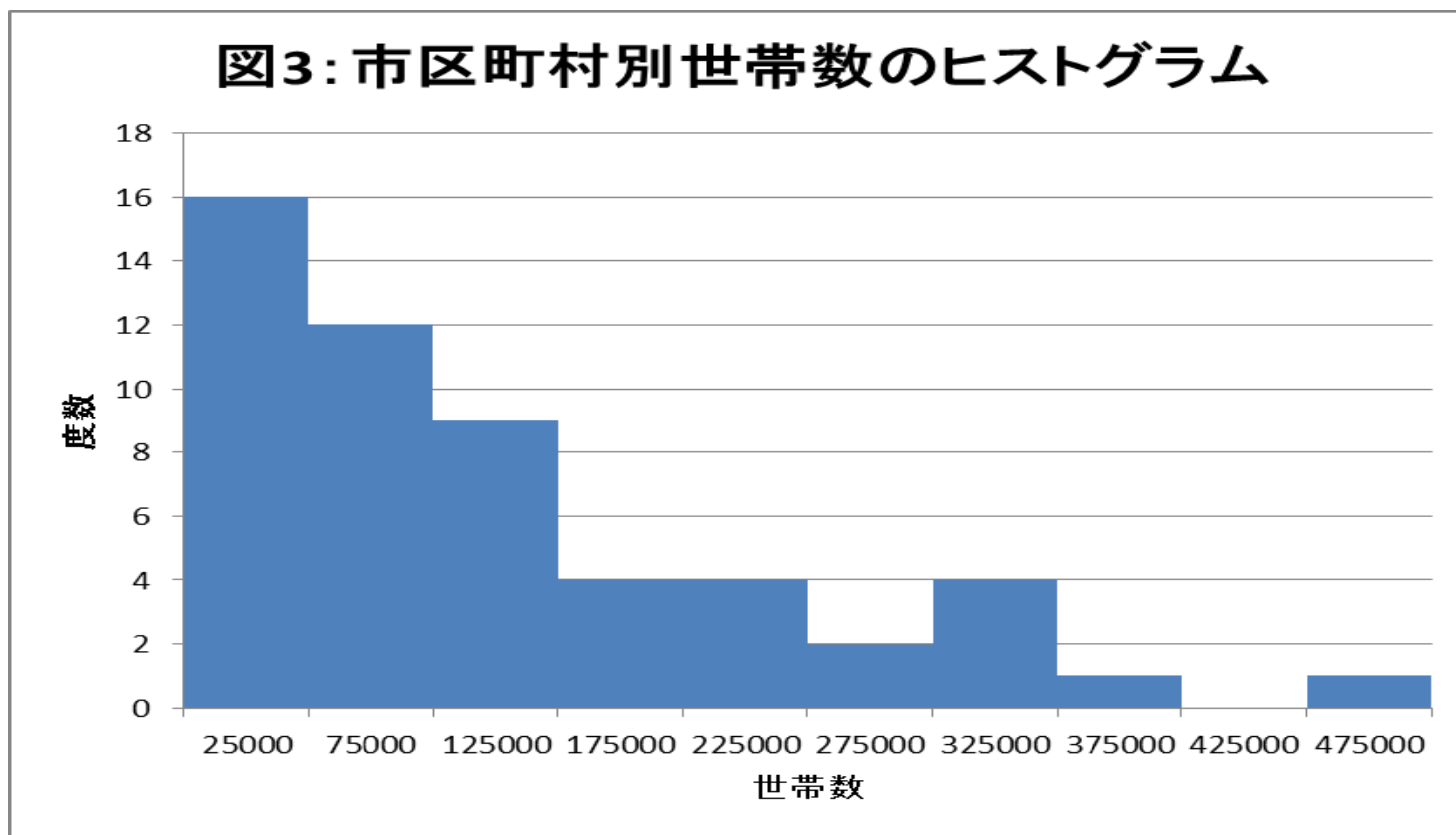
- 底辺 = 上限 – 下限 = 階級幅

- 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。

ヒストグラムの作成(2)



資料：総務省「平成27年国勢調査」人口速報集計結果

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)。

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)
 - もとも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。

ヒストグラムの見方(2)

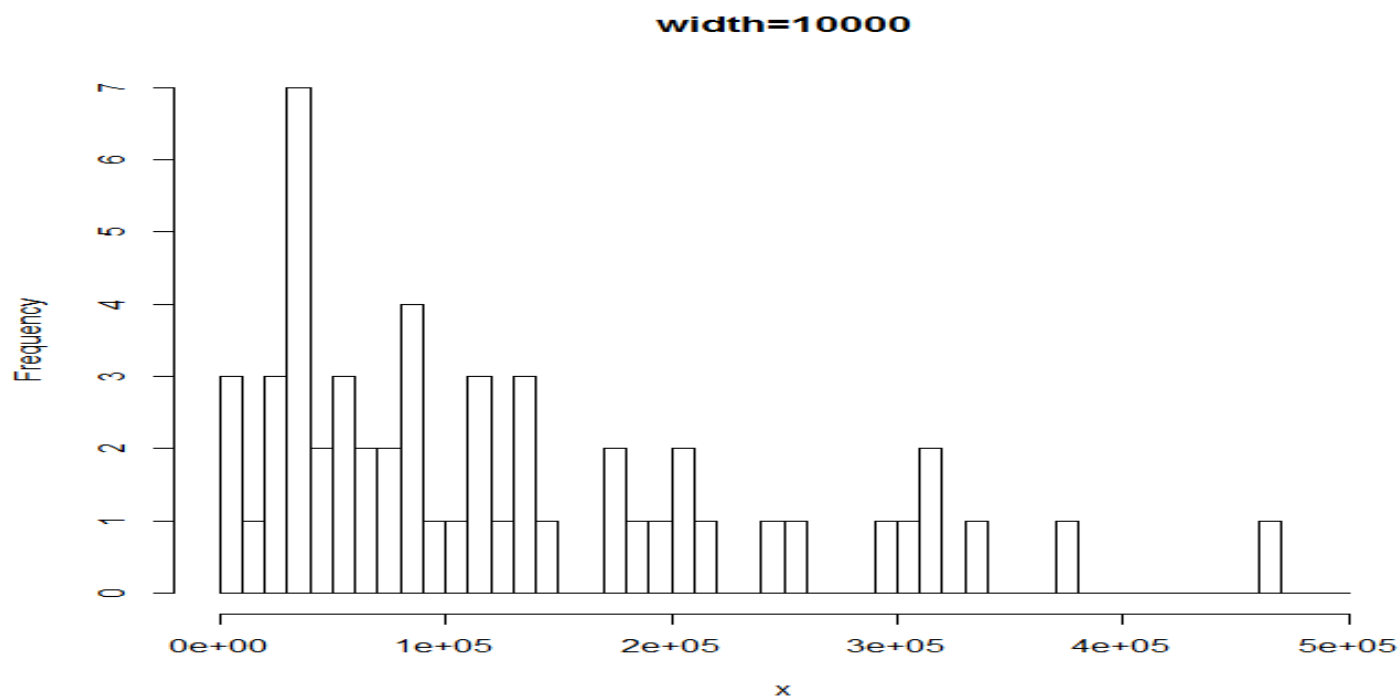
- 中心の位置
 - 分布の中心はどのあたりか。
- 散らばり具合
 - 分布の幅はどれぐらいか。
- 留意点
 - ヒストグラムの形状(見た目の印象)は階級構成に左右される。

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 $\longleftrightarrow$ 広
 - 度数分布表のふくむ情報量：大 $\longleftrightarrow$ 小
 - ヒストグラムの形状の平滑度：小 $\longleftrightarrow$ 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。

度数分布表における階級の構成(2)

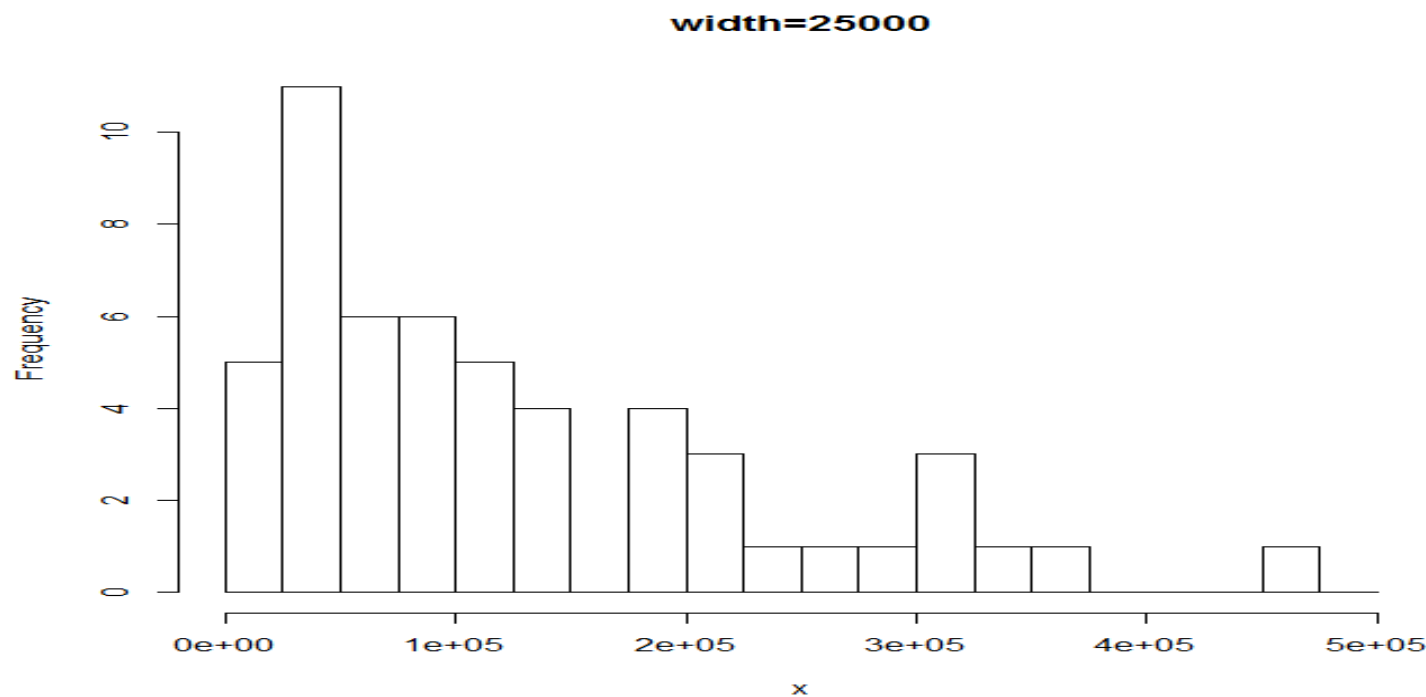
図4:ヒストグラム



資料:総務省「平成27年国勢調査」人口速報集計結果

度数分布表における階級の構成(3)

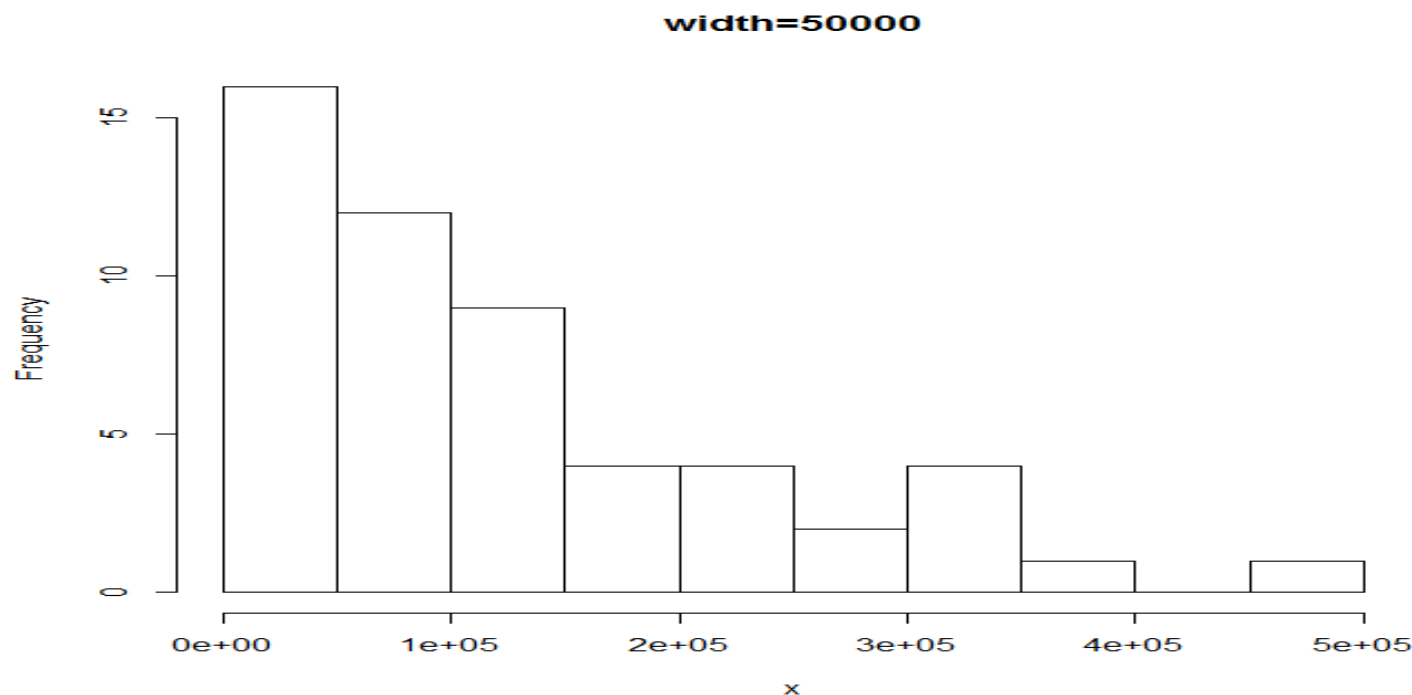
図5:ヒストグラム



資料:総務省「平成27年国勢調査」人口速報集計結果

度数分布表における階級の構成(4)

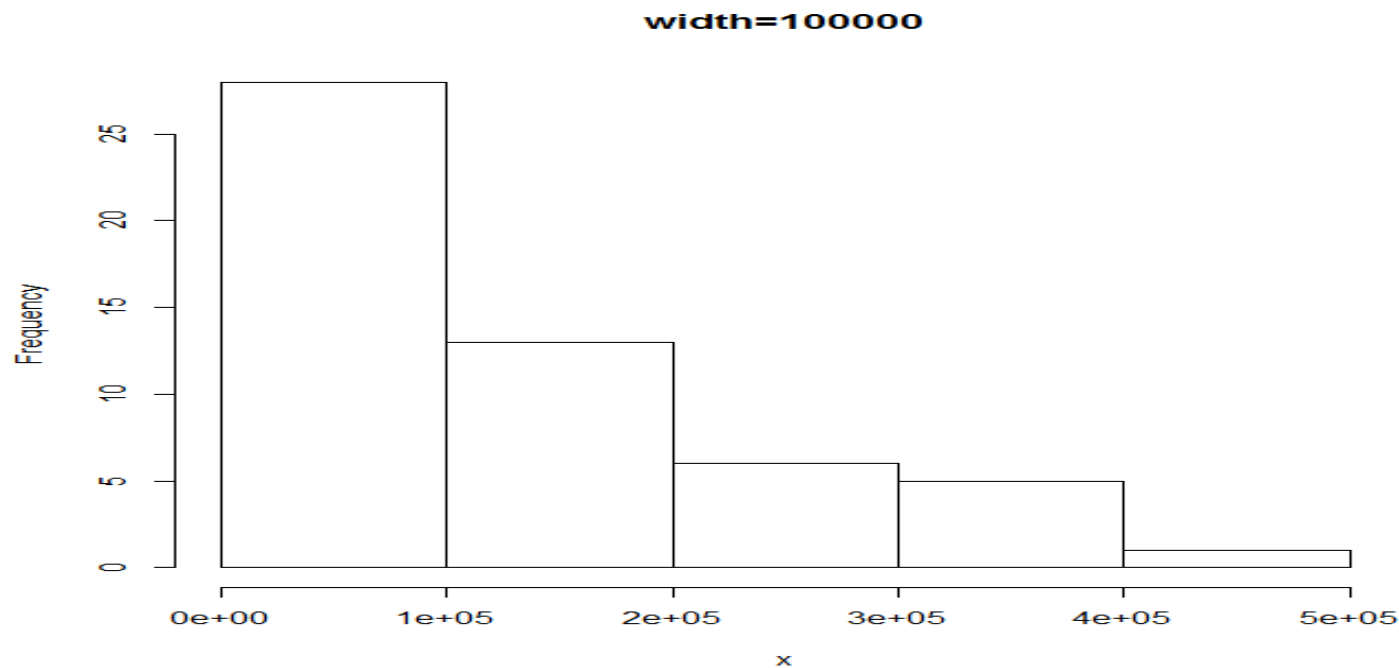
図6:ヒストグラム



資料:総務省「平成27年国勢調査」人口速報集計結果

度数分布表における階級の構成(5)

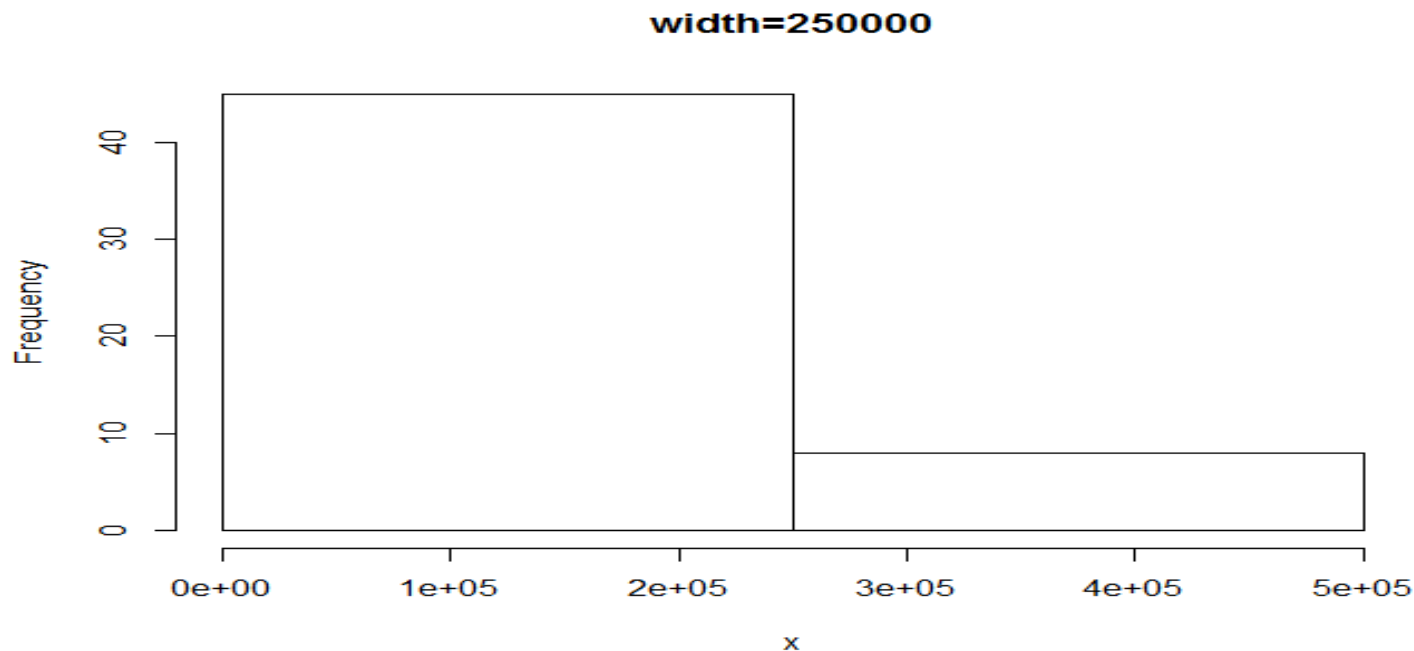
図7:ヒストグラム



資料:総務省「平成27年国勢調査」人口速報集計結果

度数分布表における階級の構成(6)

図8:ヒストグラム



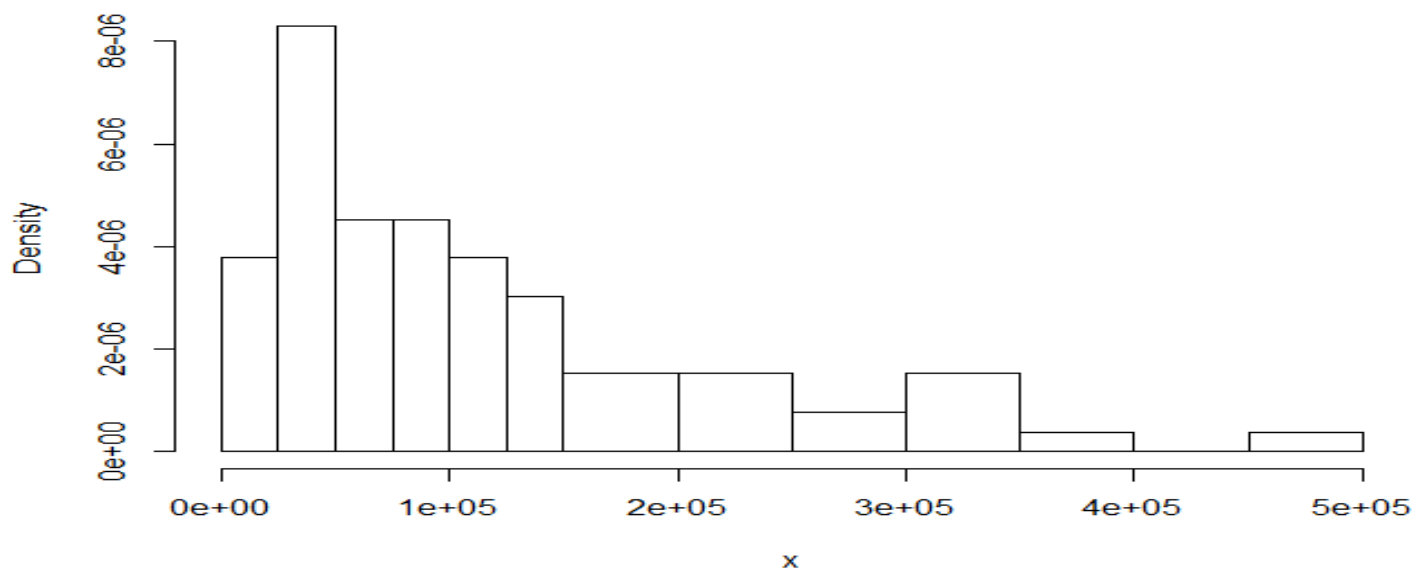
資料:総務省「平成27年国勢調査」人口速報集計結果

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: 東京都の市区町村別世帯数
 - $0 \text{ 千世帯} < x \leq 150 \text{ 千世帯}$: 階級幅 25 千世帯
 - $150 \text{ 千世帯} < x$: 階級幅 50 千世帯

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）

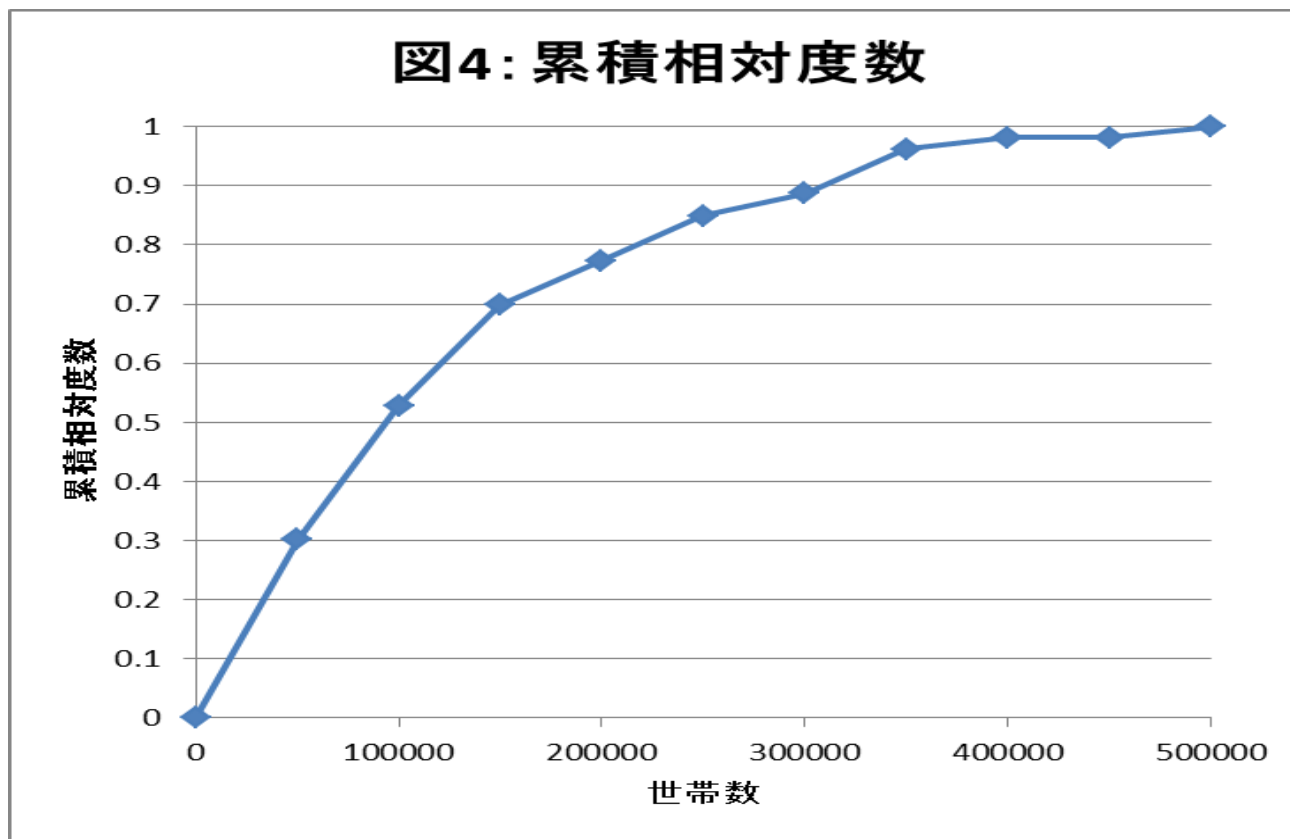


資料:総務省「平成27年国勢調査」人口速報集計結果

累積相対度数のグラフ(1)

- 累積相対度数のグラフ(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。

累積相対度数のグラフ(2)

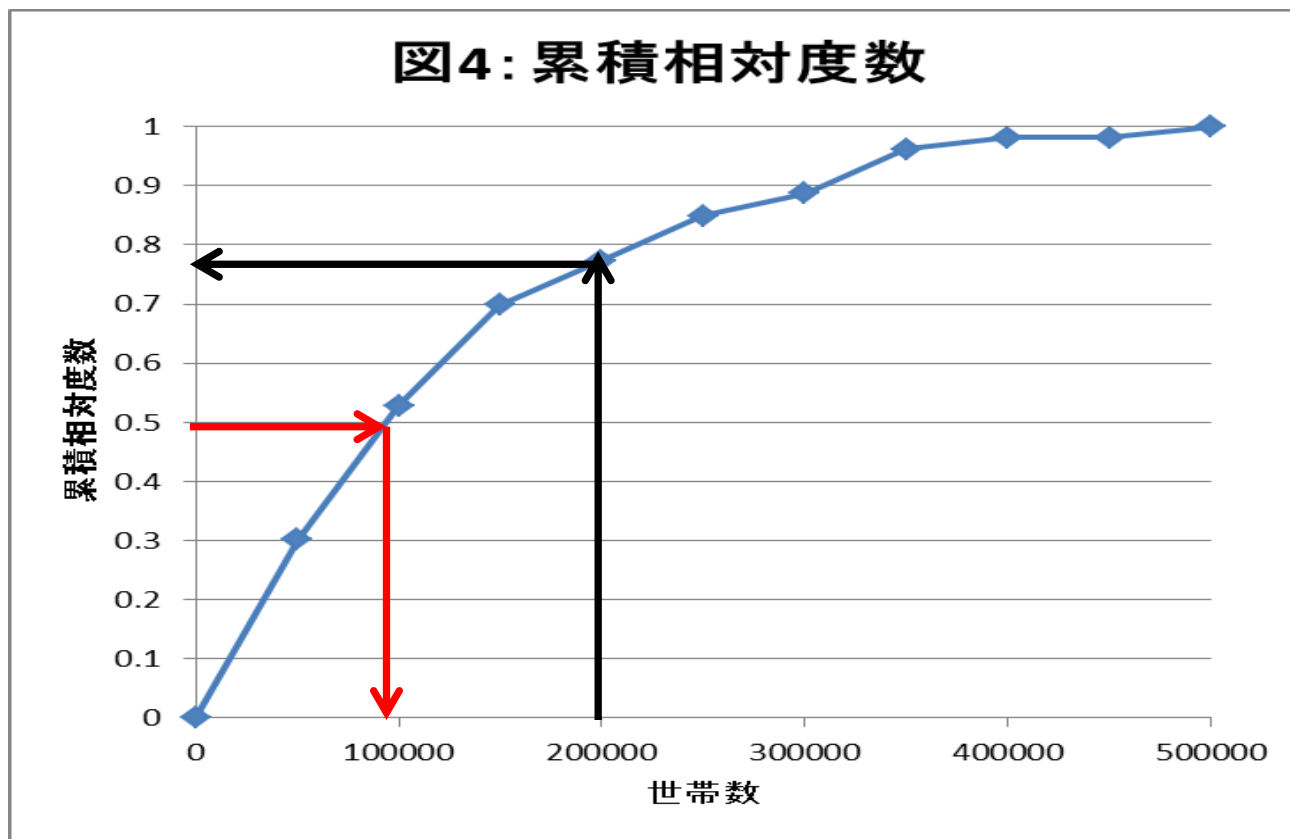


資料: 総務省「平成27年国勢調査」人口速報集計結果

累積相対度数のグラフ(3)

- 累積相対度数のグラフの利用方法
 - 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(累積相対度数)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。

累積相対度数のグラフ(4)



資料: 総務省「平成27年国勢調査」人口速報集計結果

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