



## 第4回 4月21日の授業内容

### § 2. 記述統計

- § 2.1 データ集合
- § 2.2 データの種類
- § 2.3 データの分布

4/21/09

1



## § 2. 記述統計

- 記述統計(descriptive statistics)とは  
データのかたまりを理解しやすく整理するためのテクニック・手法

4/21/09

2



## § 2.1 データ集合

### □ 定義

- 観測された個体(調査対象)のある特性(性質や状態)を数字や記号で表した「データ」のかたまり。
- データセット(data set)とも呼ぶ。

### □ データセットの性格による分類

- 「母集団(population)」 関心のある対象全体
- 「標本(sample)」 母集団の部分集合

4/21/09

3



## § 2.2 データの種類

### データの性質による分類

#### □ 量的データ

- 連続データ(計量)
- 離散データ(計数)

#### □ 質的データ

- 名義データ
- 順序データ { \* 順位  
\* 順序のあるカテゴリー

データの種類によって整理の方法も異なることに注意

4/21/09

4



## § 2.3 データの分布

### □ データセットを理解するには

- 数字の羅列では、簡単に理解できない
- まず、データを小さい順(ascending order)に並べる  
→ 分布(distribution)の作成
- データセットが大きいときには、並べるだけでは…

### □ 分布を理解しやすくする方法

- 度数(頻度)分布表(frequency distribution)
- ヒストグラム(histogram)
- 幹葉表示(stem and leaf display)

4/21/09

5



## 度数(頻度)分布表 frequency distribution

- データ(数字・記号)の羅列を表の形で整理したもの。  
変数タイプによって作成方法は異なる。  
- 度数を実際数で求める場合と、全体を1としたときのシェアで求める場合(相対度数)がある。

### □ 質的データの場合

- 質的データはとる値(ラベル・順序など)によって分類することができる。そして分類した各値に属する個体数を数えればよい。

4/21/09

7



### 度数(頻度)分布(2)

- 量的・離散データの場合
  - 質的データと同様に度数分布を作成。但し、注意すべき点は、とる値の順番に意味がある(昇順に並べる)。
  - また、ある値以下の数値をとる度数もデータの記述に意味を持つ。

→ 累積度数 cumulative frequency  
累積相対度数 cumulative relative frequency

4/21/09

8



### 度数分布(頻度分布)(3)

- 量的・連続データの度数分布
  - 連続データは実数値をとるので、離散データとは異なり、同じ値が複数観測されることは稀。



観測値を等間隔に幾つかのグループ(区間・階級)に区切り、その区間に入る個体の数をカウントする。

4/21/09

9



### 度数分布(頻度分布)(4)

- 連続データの度数分布について
    - グループ化されていることに注意。グループ化の方法によって、見え方が変わる。
  - 区間の取り方
    - 区間の個数 多すぎても少なすぎてもダメ。
- Note. スタージェス(Starjes)の公式  
区間数 =  $1 + 3.3 \times \log(\text{データセットの大きさ})$
- 区間の始点と終点

4/21/09

10



### データセット例(度数分布)

階級(区間)	階級値	度数	累積度数	相対度数	累積相対度数
0より大 ~ 10以下	5	4	4	0.21	0.21
10 ~ 20	15	9	13	0.47	0.68
20 ~ 30	25	1	14	0.05	0.74
30 ~ 40	35	1	15	0.05	0.79
40 ~ 50	45	2	17	0.11	0.89
50 ~ 60	55	2	19	0.11	1.00
計		19		1.00	

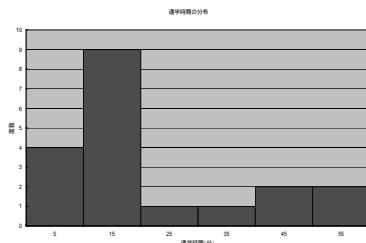
4/21/09

11



### ヒストグラム histogram

- 度数分布表、相対度数分布表を棒グラフ(bar chart)で表したもの



4/21/09

12



### 幹葉表示(stem and leaf display)

- ヒストグラムでは、区間内のデータの情報が失われてしまう。そこで、区間内のデータの情報も表示できるように工夫したヒストグラム。

- データセット例  
0 | 5000  
1 | 555555800  
2 | 5  
3 | 0  
4 | 57  
5 | 00

4/21/09

13