

## 代表値と散布度

高木 廣文  
東邦大学医学部看護学科  
国際保健看護学研究室

## データのまとめ方

### ● 質的データの場合：

- 単純集計と度数分布表の作成
- (1) カテゴリごとに人数を数える
  - (2) 人数から割合 (%) などを求める
  - (3) 表や図にまとめる

➡ 度数分布表 frequency table  
棒グラフ, 円グラフ, 帯グラフなど

## 度数分布表 frequency table

(例) ABO式血液型のデータ

血液型	度数	相対度数
A	17	42.5
B	8	20.0
O	12	30.0
AB	3	7.5
計	40	100.0

## ● 用語について

**度数**：人数, 個数, 頭数, 枚数など  
frequency

**割合**：proportion

←→昔は「比率」と誤って呼ばれていた。  
今でも, その名残がある。

## 棒グラフ bar-graph

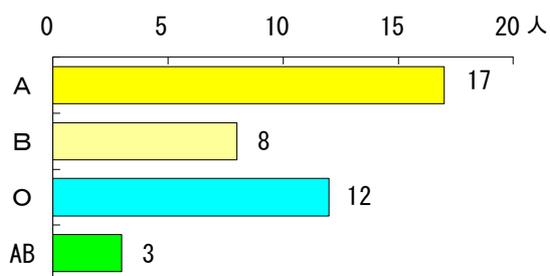


図 血液型の棒グラフ (人数)

## 円グラフ pie-graph

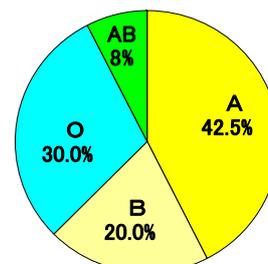
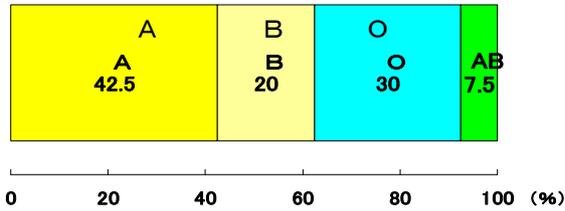


図 血液型の円グラフ

## 帯グラフ rectangular graph

図 血液型の帯グラフ



## 量的データのまとめ方

・データの分布を調べる。

度数分布表の作成。

(例) 体重のデータ:

どのように人数を数えればよいのか?

→ 5Kg ごとに幅を決めて人数を数える。

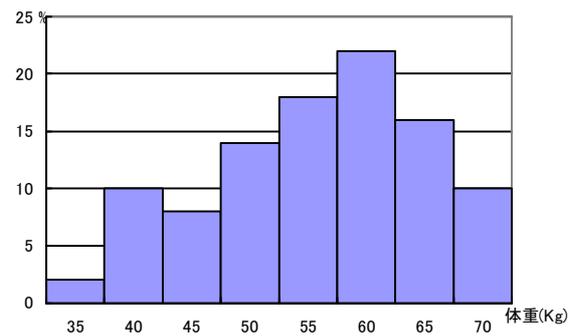
区間, 階級 class の設定。

## 量的データの度数分布

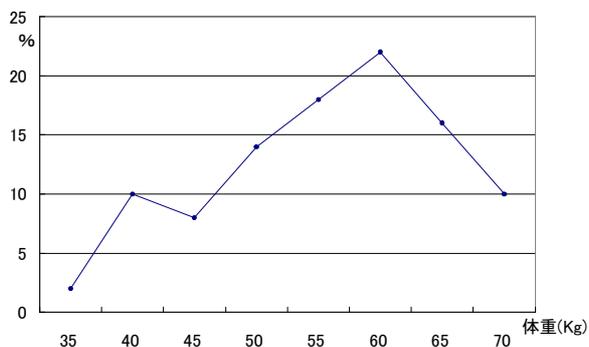
表. 体重の度数分布表

区間 (Kg)	度数 (%)	累積度数 (%)
35 ~ 40	1 ( 2.0)	1 ( 2.0)
40 ~ 45	5 ( 10.0)	6 ( 12.0)
45 ~ 50	4 ( 8.0)	10 ( 20.0)
50 ~ 55	7 ( 14.0)	17 ( 34.0)
55 ~ 60	9 ( 18.0)	26 ( 52.0)
60 ~ 65	11 ( 22.0)	37 ( 74.0)
65 ~ 70	8 ( 16.0)	45 ( 90.0)
70 ~ 75	5 ( 10.0)	50 (100.0)
計	50 (100.0)	

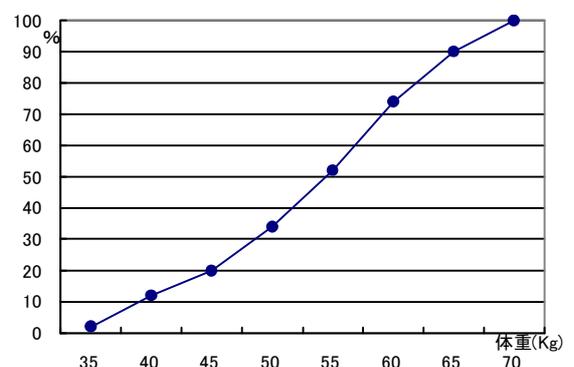
## ヒストグラム histogram



## 折れ線グラフ frequency polygon



## 累積折れ線グラフ



## 幹葉表示 stem-and-leaf display

3·   9	39
4*   02224	4042424244
4·   6788	⋮
5*   0012444	⋮
5·   566777888	⋮
6*   01111222344	⋮
6·   56677789	⋮
7*   00004	7070707074

図. 体重の幹葉表示

## 分布の代表値

### ● 分布の代表値とは

代表値 : average

→ 分布を代表する値とは何か?

- (1) 分布の真中辺のデータの値
- (2) 最も多いデータの値

→ 分布の「位置の尺度」とも呼ばれる

## 分布の3つの代表値

### ● 分布の3つの代表値

- (1) 平均値 mean
- (2) 中央値 median
- (3) 最頻値 mode

## 平均値 mean

全データの総和を標本数で割ったもの:

$$\text{平均値} = \frac{\text{データの合計}}{\text{標本数}}$$

記号 : n 個のデータを  $x_1, x_2, \dots, x_n$

$$\text{平均値} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

## 平均値の例:

例) 5人の体重のデータ : 50, 45, 60, 70, 55Kg

$$\begin{aligned} \text{平均値} &= \frac{50+45+60+70+55}{5} \\ &= 280 \div 5 = 56 \text{ (Kg)} \end{aligned}$$

## 中央値 median

データを大きさの順に並べた場合, ちょうど真ん中の順位にくるデータのもつ値。

N個のデータ大きさの順に並べる:

$x_1 \geq x_2 \geq \dots \geq x_N$   
とすると,

● Nが奇数 : 中央値  $x_{Med} = x_{\left(\frac{n+1}{2}\right)}$

● Nが偶数 : 中央値  $x_{Med} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$

## 中央値の例1:

例) 5人の体重のデータ : 50, 45, 60, 70, 55Kg

(1) まず大きさの順に並べ替える :

→ 45, 50, **55**, 60, 70 (Kg)

(2) 標本数5は奇数なので,

$(5+1)/2 = 3$  番目

のデータが中央値 → 中央値 = 55 (Kg)

## 中央値の例2:

6人の体重のデータ : 50, 45, 65, 60, 70, 55Kg

(1) まず大きさの順に並べ替える :

→ 45, 50, **55, 60**, 65, 70 (Kg)

(2) 標本数6は偶数なので,  $6/2 = 3$  番目

と4番目のデータの平均値が中央値 :

→ 中央値 =  $(55+60)/2 = 57.5$  (Kg)

## 最頻値 mode

最も人数(度数)の多いデータのもつ値。  
実際には、標本数が少ない場合、データが連続的なことから、各データの人数は少なくなり、どのデータが最頻値かを定めるのは困難。



度数分布表の利用

最も度数の多い区間の真中の値(級心)を最頻値とする。

## その他の位置の尺度

● 最小値 minimum value : データ中最小の値  
● 最大値 maximum value : データ中最大の値

● パーセンタイル percentile (百分位) :  
大きさの順にデータを並べ、小さい方から累積して何パーセントの点にあるかを示す。

→ 5, 10, **25, 50, 75**, 90, 95パーセンタイル  
(第1, 2, 3四分位 quartile)

## 分布の散布度

各データは異なった値を持つので、その分布には広がりがある。  
そのばらつき具合、代表値からの平均的な散らばり具合を示す。

- 1) 分散 variance, Var
- 2) 標準偏差 standard deviation, SD
- 3) 変動係数 coefficient of variation, CV
- 4) 範囲 range, R
- 5) 平均偏差 mean deviation

## 偏差について

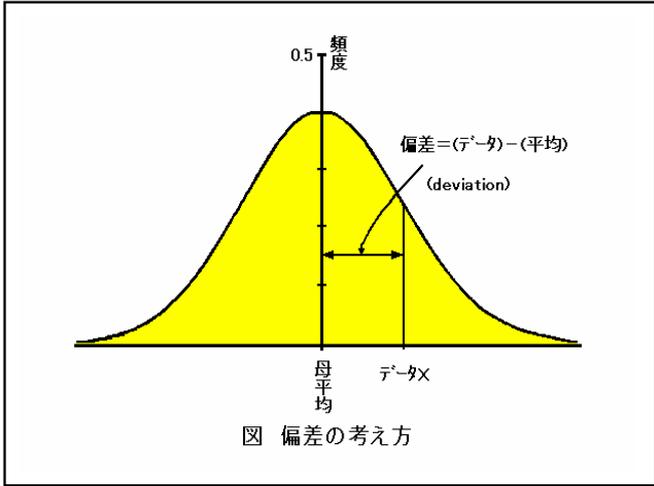
代表値とデータとの差、普通は代表値として平均値を用いる。

偏差 deviation = [データ] - 平均

(例) 身長が180cmの場合、平均身長が170cm

身長の偏差 =  $180 - 170 = 10$  (cm)

24



### 偏差に基づく散布度

分布の散布度をどのようにして求めればよいか

- 偏差の平均値は？
- 偏差の合計は常に0。使用不可。  
(練習問題1:確かめてみよう)

---

偏差に正負があるので、全て正にすればよい。

- 偏差の絶対値, 偏差の2乗(平方)

26

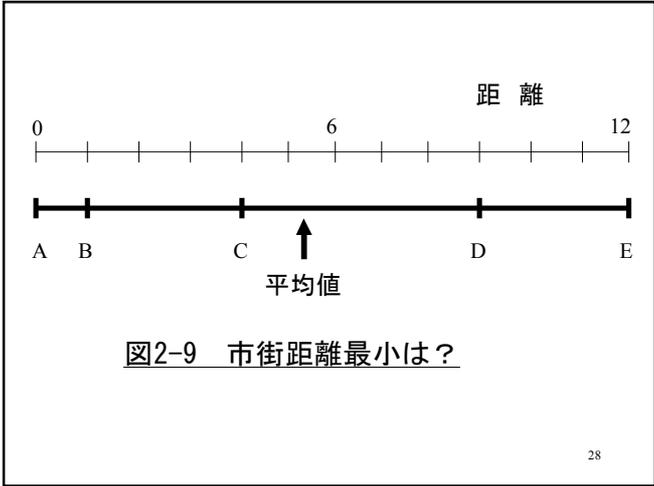
### 平均値からの平均偏差

◎ 平均値からの偏差の絶対値の平均値:  
n個のデータを  $x_1, x_2, \dots, x_n$

$$\text{平均偏差} = \frac{\text{各ケースの偏差の絶対値の合計}}{\text{標本数}} = \frac{1}{n} \sum_{i=1}^n |x_i - x_{Med}|$$

- 統計的な扱いが難しいため、実際には、ほとんど使用されない。

27



### 中央値からの平均偏差(市街距離最小)

どこにコンビニを作れば、各人からの距離の総計が最小になるか？

- ↓ 平均値=5.2: 総距離=5.2+4.2+1.2+3.8+6.8=21.2
- ↓ 中央値=4 : 総距離=4+3+0+5+8=20

29

### 分散 variance

◎ 平均値からの偏差の平均平方和:  
n個のデータを  $x_1, x_2, \dots, x_n$

$$\text{分散 } s^2 = \frac{\text{偏差の2乗の合計}}{\text{標本数}} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

30

## 分散の例:

◎ 5人の体重のデータ : 50, 45, 60, 70, 55Kg

平均値 = 56Kg

$$\begin{aligned} \text{分散} &= [(50-56)^2 + (45-56)^2 + (60-56)^2 \\ &\quad + (70-56)^2 + (55-56)^2] \div 5 \\ &= [36 + 121 + 16 + 196 + 1] \div 5 \\ &= 370 \div 5 \\ &= 74 \text{ (Kg}^2\text{)} \end{aligned}$$

31

## 標準偏差 standard deviation (SD)

分散は偏差の2乗の合計から計算

→ 単位も2乗 : 体重 = Kg<sup>2</sup>, 身長 = cm<sup>2</sup>, etc

● 分散の平方根を計算し, 単位を戻す。

$$\text{SD} = \sqrt{\text{分散}}$$

32

## 標準偏差の例:

◎ 5人の体重のデータ : 50, 45, 60, 70, 55 Kg

平均値 = 56 (Kg)

分散 = 74 (Kg<sup>2</sup>)



$$\text{標準偏差 } S = \sqrt{74} = 8.60$$

33

## 変動係数 CV(Coefficient of Variation)

50人の身長の標準偏差は5 cm, 体重の標準偏差は5 kgであった。

Q. 身長と体重のばらつき具合はどちらが大きいのか, それとも等しいのか?

単位が異なるので比較できない!  
単位をそろえる必要がある。

34

## 変動係数の定義

$$\text{変動係数} = \frac{\text{標準偏差}}{\text{平均値}} \times 100$$

平均値を100としたときの標準偏差の大きさの程度を示す。

35

## 変動係数の計算例:

◎ 5人の体重のデータ : 50, 45, 60, 70, 55 Kg

平均値 = 56 (Kg)

分散 = 74 (Kg<sup>2</sup>)

標準偏差 = 8.602 (Kg)



$$\text{変動係数} = \frac{8.602}{56} \times 100 = 15.36$$

36

## ● 散布度に関するその他の話題

### ▼ 偏差値とは何か？

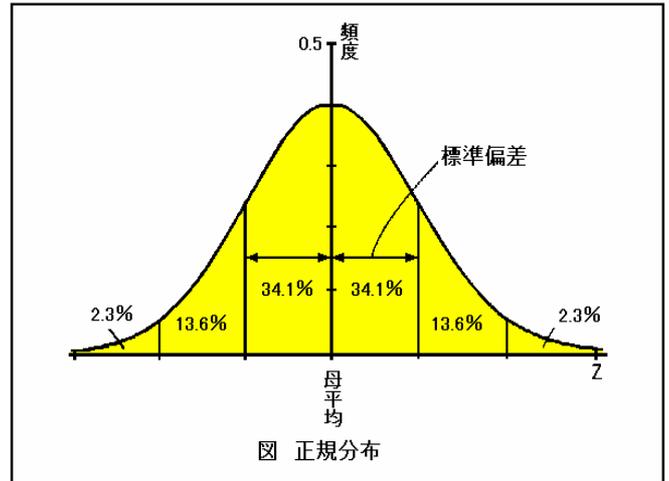
偏差は平均値からの差。

→ データの標準化 standardization

### 標準化：

データの平均が0，分散が1になるようにデータを変換すること。

37



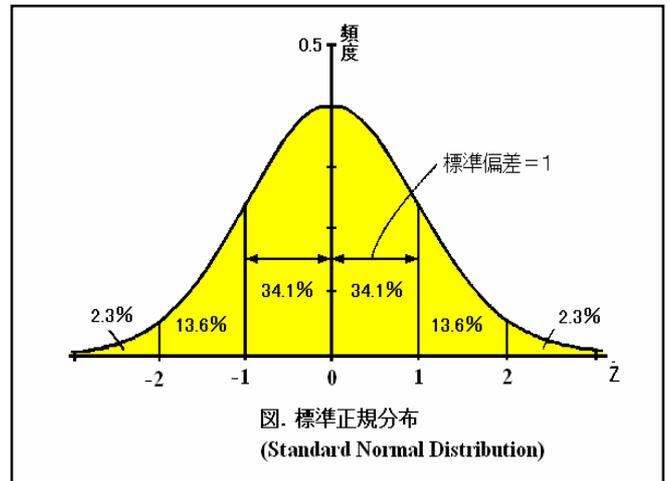
## ● データの標準化と偏差値

平均が $\mu$ （ミュウ），標準偏差が $\sigma$ （シグマ）の場合，あるデータ $x$ を，

$$z = \frac{x - \mu}{\sigma} \quad : \text{ (平均0, 分散1) }$$

偏差値  $T = 10 \times z + 50$   
平均50，標準偏差10

39



## 偏差値の計算例：

統計学の平均値が75点，標準偏差10点の場合：

A君90点，B君70点の偏差値は？

$$\text{A君の偏差値} = \frac{90 - 75}{10} \times 10 + 50 = 65$$

$$\text{B君の偏差値} = \frac{70 - 75}{10} \times 10 + 50 = 45$$

41

## ● 四分位間範囲と四分位偏差

データの順位に基づく散布度

四分位間範囲 = 第3四分位 - 第1四分位  
interquartile range

四分位偏差 = 四分位間範囲 / 2  
quartile deviation

42