

5. 割合, 率, 比

高木 廣文
東邦大学看護学部
国際保健看護学研究室

1

割合, 率, 比の違いは?

1. 割合 proportion
2. 率 rate
3. 比 ratio

2

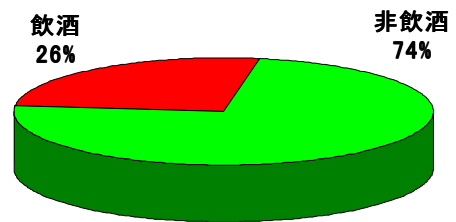
1. 割合 proportion

(例) 主婦300人中 1週間に2回以上飲酒すると答えたものは78人であった。

$$\text{飲酒者の割合} = \frac{78}{300} = 0.26$$

3

割合の意味



● 全体に対する特定部分の占める大きさ

4

2. 率 rate

(例) ある町の40歳以上の1万人を10年間追跡調査したところ, 肺癌で50人が死亡した。

◆ 観察人年 (人-時間 person-time)

全ての観測された対象者と観測時間の積和

$$\text{発生率} = \frac{\text{事象発生数}}{\text{観測人-時間の総計}} = \frac{50人}{10万人年}$$

5

率の単位は時間!

(例) ある町の40歳以上の1万人を10年間追跡調査したところ, 5千人は5年間, 残りの5千人は10年間観測できた。全体では肺癌で50人が死亡した。

$$\begin{aligned} \text{発生率} &= \frac{50}{5000 \times 5 + 5000 \times 10} = \frac{50人}{75000人年} \\ &= 66.7 / 10万年 \end{aligned}$$

6

3. 比 ratio

(例) 主婦300人中 1週間に2回以上飲酒すると答えたものは78人であった。飲酒者と非飲酒者の割合の比 R は？

$$R = \frac{\text{飲酒者数}}{\text{非飲酒者数}} = \frac{\text{飲酒割合}}{\text{非飲酒割合}}$$

$$= \frac{78}{222} = \frac{(78/300)}{(222/300)} = 0.351$$

7

率の比 (rate ratio)

(例) 喫煙者と非喫煙者の肺癌死亡率が10万人年あたり400人と40人とすると、両者の死亡率の比は？

$$RR = \frac{\text{喫煙者死亡率}}{\text{非喫煙者死亡率}} = \frac{400}{40} = 10$$

喫煙の肺癌死亡率への影響の大きさを示す。

⇒ **レイト比, 率比 rate ratio**

8

リスクの評価と疫学研究

喫煙者と非喫煙者の死亡率の**レイト比**は、「**喫煙の影響**」を評価するために使用できる。

⇒ 喫煙の肺癌死亡に対する **リスク risk** の評価。
「**疫学研究**」で多用される。

9

疫学研究と効果の指標

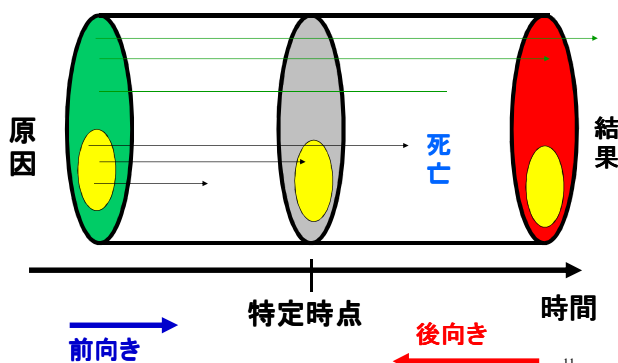
疫学研究は基本的に**非実験・観察研究**：

- ・人口集団研究
- ・**実験的研究** → **介入研究：臨床研究**

調査方法によって、リスク評価のための効果の指標、統計的方法が異なる。

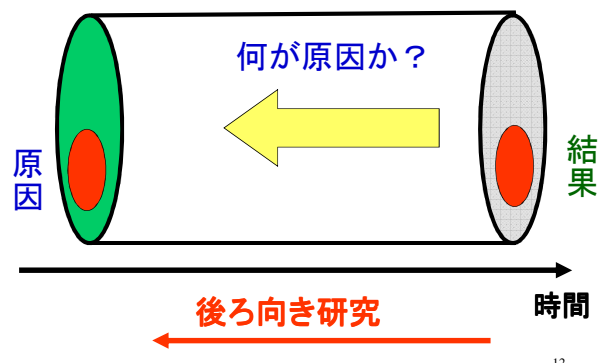
10

時系列調査法の模式図



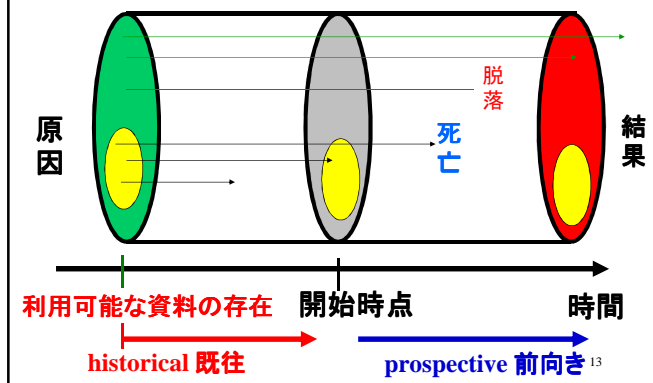
11

ケース・コントロール研究の概念



12

コホート研究の分類



断面研究での効果の指標

特定要因の有無と疾患等の有無の関係について分析可能。

【用語】

- ・ 曝露グループ exposed group, exposure
- ・ 非曝露グループ nonexposed group, nonexposure
- ・ 有病割合 prevalence (proportion) (有病率)

14

有病割合とは？

表. 断面研究での四分表

要因	疾患		計
	あり	なし	
あり	a	b	x
なし	c	d	y
計	m	n	N

- 曝露グループ : 有病割合 $p_1 = a/x$
- 非曝露グループ : 有病割合 $p_0 = c/y$

15

リスク差とリスク比の定義

- ・ リスク差 risk difference, RD

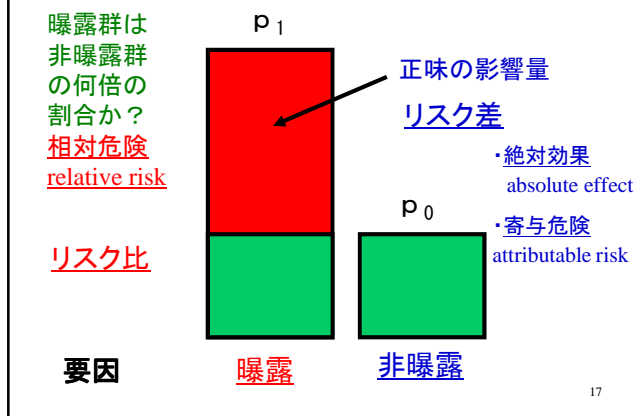
$$RD = \text{曝露群の有病割合} - \text{非曝露群の有病割合} = p_1 - p_0 = a/x - c/y$$

- ・ リスク比 risk ratio, RR

$$RR = \frac{\text{曝露群の有病割合}}{\text{非曝露群の有病割合}} = \frac{p_1}{p_0} = \frac{a/y}{c/x}$$

16

リスク比とリスク差の意味



17

リスク差とリスク比の例

表. 飲酒習慣と高血圧の既往

飲酒	高血圧の既往		計
	あり	なし	
あり	8	70	78
なし	41	181	222
計	49	251	300

・ 飲酒群 :

$$p_1 = 8/78 = 0.103$$

・ 非飲酒群 :

$$p_0 = 41/222 = 0.185$$

- ・ リスク差 $RD = p_1 - p_0 = -0.082$
- ・ リスク比 $RR = p_1 / p_0 = 0.555$

18

コホート研究の効果の指標

特定人口集団を長期間追跡するために、疾患の発生、死亡などが観測できる。

- ・ **発生率** incidence rate, IR
観測された総人-時間で、特定事象の発生数を除したもの。

$$IR = \frac{\text{事象発生数}}{\text{観測人-時間の総計}}$$

19

レイト差とレイト比

- ・ **曝露群** : 発生率 IR_1
- ・ **非曝露群** : 発生率 IR_0
- ・ **レイト差** rate difference, RD (率差)

$$RD = IR_1 - IR_0$$

- ・ **レイト比** rate ratio, RR (率比)

$$RR = IR_1 / IR_0$$

20

レイト差とレイト比の例

表. 喫煙と肺癌死亡率 (10万人年あたり)

年齢(歳)	非喫煙	10~20本	21~39本
55~64	40	250	400
65~74	80	500	720

- ・ 55~64歳で10~20本/日の喫煙者は非喫煙者に比べて,
レイト差 $RD = 250 - 40 = 210$ (人/10万人年)
レイト比 $RR = 250 / 40 = 6.25$

21

レイト差とレイト比の例(続)

表. 喫煙の肺癌死亡率に対するレイト差とレイト比

年齢(歳)	喫煙本数 (/ 日)	
	10~20本	21~39本
レイト差 RD :		
55~64	210	360
65~74	420	640
レイト比 RR :		
55~64	6.25	10.0
65~74	6.25	9.0

22

ケース・コントロール研究での効果の指標

- ・ **患者-対照研究**
人的、経費的に負担が少なく、病院などで研究を行う場合、特定患者グループに対して、適当なコントロール・グループを設定し、比較研究を行う。

- ・ **コントロール** control (対照)
ケースの母集団での曝露分布の情報を与えることができる曝露状況をもつ集団。

23

ケース・コントロール研究での四分表

	要 因		計	
	あり	なし		
患者群	a	b	x	人数が固定されている
対照群	c	d	y	
計	m	n	N	

- 要因の有無ごとに有病割合、疾患の発生率などを求めることはできない。

24

オッズとオッズ比

・ 曝露オッズ exposure odds :

曝露割合を非曝露割合で除したもの。

患者群での曝露オッズ = a/b

対照群での曝露オッズ = c/d

・ オッズ比 odds ratio, OR :

相対リスク relative risk の推定値 :

$$\text{オッズ比OR} = \frac{\text{患者群での曝露オッズ}}{\text{対照群での曝露オッズ}} = \frac{a d}{b c}$$

25

オッズ比の計算例

表. 膀胱癌と喫煙の関係

	喫煙習慣		計
	あり	なし	
患者群	284	69	353
対照群	261	114	375

$$\text{オッズ比} = \frac{284 \times 114}{69 \times 261} = 1.798$$

26

6. 統計学の基礎

高木 廣文
東邦大学看護学部
国際保健看護学研究室

e-mail: halwin@med.toho-u.ac.jp

<http://homepage2.nifty.com/halwin/takagi.html>₂₇

統計学と統計について

統計学 statistics とは何か?

- ・ 統計 : 統計をとる (?)
- ・ 統計学 : 統計学を使う (?)

28

「統計をとる」とは?

- ・ アンケート調査で学生のアルバイト実施を調べる。
- ・ ある病院の診療科別外来患者数を調べる。

⇒ データを収集する (データをとる)。
集計をする。人数を数える。等々。

29

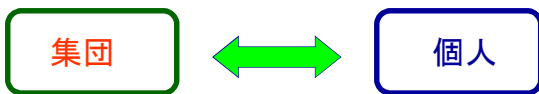
統計学とは?

- ・ 数理統計学
- ・ 生物統計学
- ・ 経済統計学
- ・ 看護統計学
- ・

● データに内在する傾向を明らかにするための科学的な方法論を与える

30

統計学の対象は何か？



集団がもつ各項目や特性などの傾向についてデータから検討するための方法を提供

31

統計学の立場

●統計学の特徴

- ・現象の数量化（データ化）
 - ・各種検査値
 - ・臨床的な症状や患者の性格特性
 - ・QOLなど
- ・「質」を「量」として把握
 - ・客観性を高める。
「再現性」, 「比較可能性」

科学的な研究に不可欠

32

統計学を使う目的

統計学を目的により大きく2つに分けて考えることがある。

(1) 特定の事象の「記述」

「記述統計学」 descriptive statistics

(2) 調査結果や研究結果の「一般化」

「推測統計学」 inferential statistics

検定, 推定

33

記述統計学の方法

データをまとめて, ある特性を示す。

●図や表を用いて示す方法

●ひとつの数値で示す方法

34

データを図示する手法

「円グラフ」, 「帯グラフ」, 「棒グラフ」, 「ヒストグラム」, 「折れ線グラフ」, 「幹葉表示 stem-and leaf display」, 「箱ヒゲ図 box and whisker plot」, 「相関図（散布図）」など

35

データを要約するための指標

「代表値」— 平均値, 中央値, 最頻値

「散布度」— 分散, 標準偏差, 変動係数

「相関係数」, 「割合」, 「クロス表」など

36

推測統計学の方法

データから得た結果を一般化，普遍化する。

●推定 estimation

データ（**標本**）から一般集団（**母集団**）の特性を求める：

⇒**母平均値**，**母比率の信頼区間**など

●検定 test

データから一般集団の特性に関する仮説を検証 ⇒**独立性の検定**，**無相関の検定**など

37

データ解析について

従来の多くの研究：

統計的検定の多用— **確証的解析**
confirmative analysis

記述的方法を多用，データに依存した解析：
探索的データ解析 exploratory data analysis

幅広い知識や教養，また洞察力・直観力も極めて重要。方法の選択や結果の解釈を正しく行うには，統計学に関する正確な知識も必要

38

データとは何か

例 1. A子さんの身長は160cmです。

- ・ A子さんの身長の測定結果。
- ・ A子さんの身長の「**データ**data」。

例 2. 学校で行われる身体計測

- ・ あるクラス30人の身長の一覧表。
- ・ そのクラスの「**身長**の**データ**」。

39

ラベリングについて

例 3. A子さんの血液型はA型です。

- ・ 血液型のデータを示している点は，身長の場合と全く同じである。
- ・ 血液型は，ある特定の反応の有無により A, B, O, AB の4タイプに分類

⇒**標識付け**

（ラベリング labeling, ラベル付け）

40

データの定義

個体のある特性について測定を行い，
適当なラベルを付けたもの。
もしくは，その全体，およびそれらをまとめたもの

41

データの分類

1) 身長や体重のデータ

- ・ ある「物差し」を用いて測定：数値で表現
 - ・ データを足したり引いたりできる。
- ⇒**量的データ** quantitative data

2) 血液型のデータ

- ・ ある特性に名称をつけたもの。
 - ・ それぞれを足したりすることは不可能。
- ⇒**質的データ** qualitative data

42

量的データの分類

例 1. A 子さんの体重は50Kg,
K 子さんの体重は60Kgです。

- (1) K 子さんは A 子さんより10Kg体重が重い。
 - ・データの「差」の計算ができる。
 - (2) K 子さんは A 子さんの1.2倍の体重がある。
 - ・データの「比」の計算ができる。
- ⇒ 「比尺度 ratio scale」によるデータ

43

間隔尺度

例 2. 一日の最高気温20°C, 最低気温10°C。

- (1) 一日の気温の差は10°Cである。
⇒ 差の計算可能である。
- (2) 最高気温は最低気温の2倍である？
⇒ × 比の計算不可。

「間隔尺度 interval scale」によるデータ

原点0のもつ意味による相違：

- ・負のデータの存在。
- ・比尺度に負のデータはない！

44

質的データの分類

例 1. 血液型のデータの場合：

A 子さんはA型, K 子さんはAB型。

- ・差の計算： $A - AB = A(1 - B)$ ？
- ・比の計算： $AB / A = B$ ？
- ・A型とB型の差や比を取ることは不可能。
どちらが大きいともいえない。

⇒ 「名義尺度 nominal scale」によるデータ

45

順序尺度

例 2. 「あなたは寝起きはよいですか」
のような質問項目への回答

1. 非常によい 2. よい 3. 悪い 4. 非常に悪い

- ・各カテゴリについての数値1~4の差や比は計算できない。
- ・数値が大きくなるにつれ、寝起きが悪くなるという、順序がある。

⇒ 「順序尺度 ordinal scale」によるデータ

46

データの分類再考

●データの測定尺度によるまとめ

- (1) 量的データ：
 - (A) 比尺度によるデータ
 - (B) 間隔尺度によるデータ
- (2) 質的データ：
 - (C) 順序尺度によるデータ
 - (D) 名義尺度によるデータ



47

データの「質」と基本的統計手法

(1) 量的データ：

平均値, 分散, 標準偏差, 相関係数, など

(2) 質的データ：

人数, 割合, クロス表など

●求められる基本統計量が異なる！

48

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

49

度数分布表 frequency table

(例) AB0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

50

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。

51

棒グラフ bar-graph

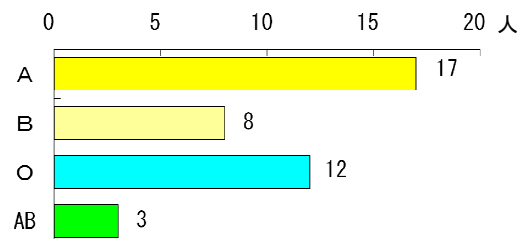


図 血液型の棒グラフ (人数)

52

円グラフ pie-graph

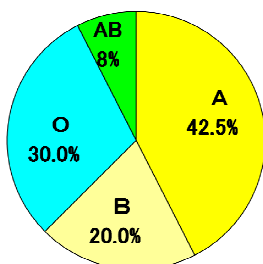


図 血液型の円グラフ

53

立体円グラフ

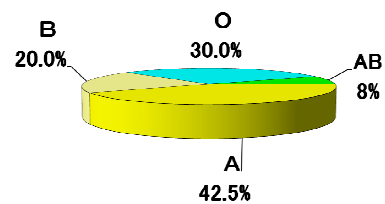


図 血液型の円グラフ

54

帯グラフ rectangular graph

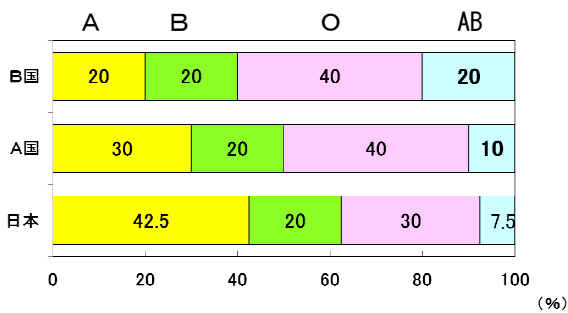


図 血液型の帯グラフ

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？

→ 5 Kg ごとに幅を決めて人数を数える。

区間, 階級 class の設定。

56

量的データの度数分布

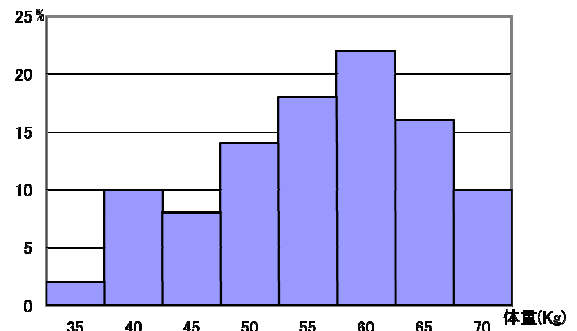
表. 体重の度数分布表

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)	

(注) 区間の左側は以上、右は未満を示す。

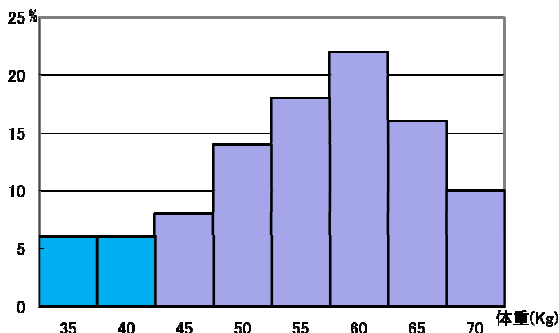
57

ヒストグラム histogram



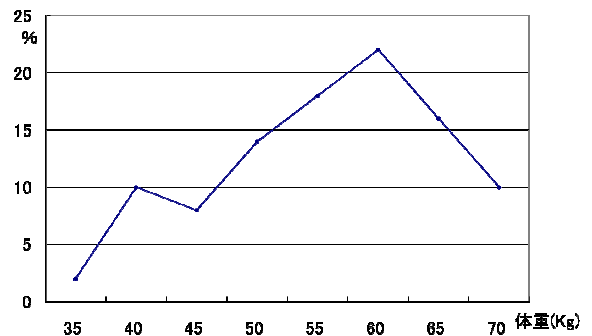
58

ヒストグラム histogram 2



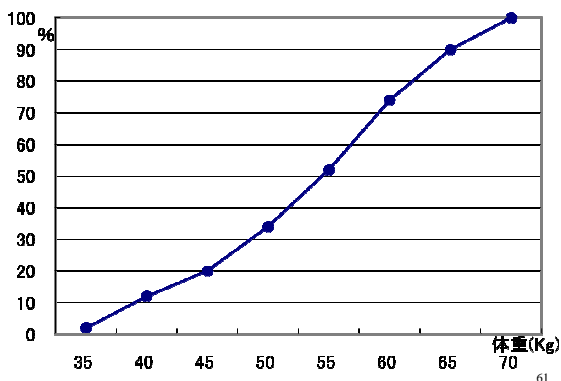
59

折れ線グラフ frequency polygon



60

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

62

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

63

分布の3つの代表値

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

64

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

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

65

平均値の例 :

例) 5人の体重のデータ : 50, 45, 60, 70, 55Kg

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

66

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：

$$x_1 \geq x_2 \geq \dots \geq x_n$$

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

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

67

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)

68

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)

69

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決めるのは困難。



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。

70

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)

71

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

72

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)

73

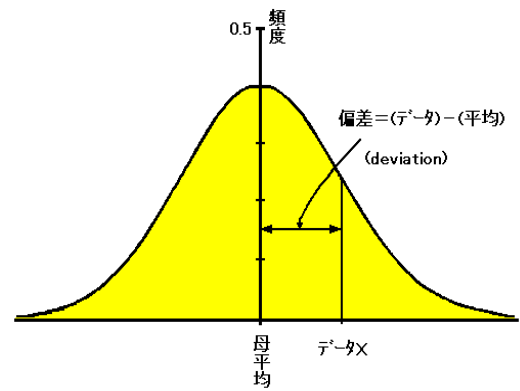


図 偏差の考え方

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。

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）

75

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

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

● 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。

76

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

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

77

分散の例：

◎ 5人の体重のデータ：50, 45, 60, 70, 55Kg

平均値 = 56Kg

$$\text{分散} = [(50-56)^2 + (45-56)^2 + (60-56)^2 + (70-56)^2 + (55-56)^2] \div 5$$

$$= [36 + 121 + 16 + 196 + 1] \div 5$$

$$= 370 \div 5$$

$$= 74 \text{ (Kg}^2\text{)}$$

78

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=Kg²，身長=cm²，etc

●分散の平方根を計算し，単位を戻す。

$$SD = \sqrt{\text{分散}}$$

79

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 (Kg²)



標準偏差 $S = \sqrt{74} = 8.60$

80

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。

81

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。

82

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 (Kg²)

標準偏差= 8.602 (Kg)



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

83

散布度に関するその他の話題

▼ 偏差値とは何か？

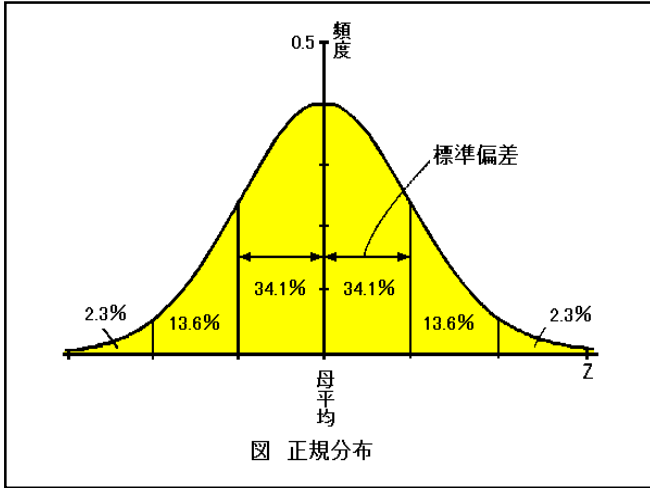
偏差は平均値からの差。

→ 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。

84



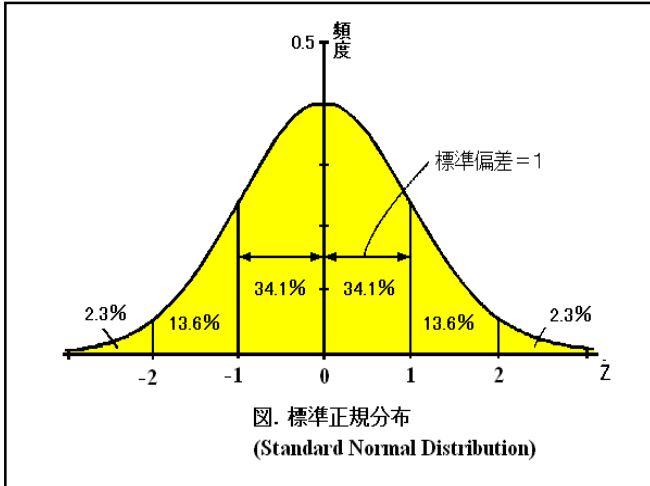
データの標準化と偏差値

平均が μ (ミュウ), 標準偏差が σ (シグマ) の場合, あるデータ x を,

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

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

86



偏差値の計算例:

統計学の平均値が75点, 標準偏差15点の場合:
 A君90点, B君60点の偏差値は?

$$\text{A君の偏差値} = \frac{90 - 75}{15} \times 10 + 50 \doteq 60$$

$$\text{B君の偏差値} = \frac{60 - 75}{15} \times 10 + 50 \doteq 40$$

88