

2017年度日本疫学会スライドコンテスト受賞作品

基本的な統計 —データの記述—

藤田保健衛生大学医学部

公衆衛生学

柿崎 真沙子

基本的な統計

- ① データの種類
- ② データの要約値
- ③ データの分布を見る
- ④ 割合・比・率

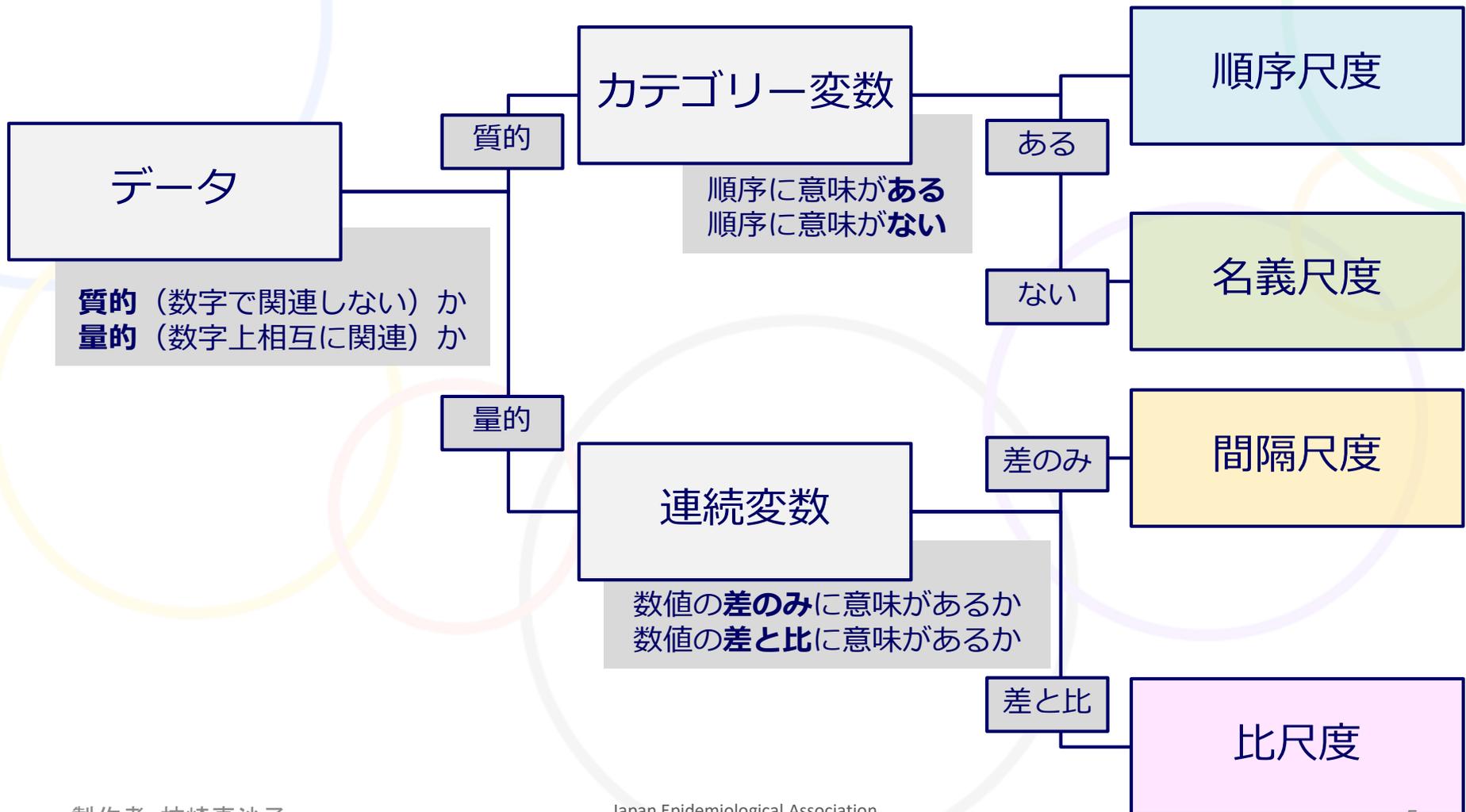
統計学の基本

- 数値や図表で分かりやすく示す
- 難しい統計テクニック（分散分析、重回帰分析等々の**分析統計**）よりも単なる平均値や割合などの方が重要なことも
- 平均値や割合→**記述統計**

データの種類を把握する

- データの種類によって、それ以降の統計処理が全く異なる
- 統計ソフトを利用する場合、「データの種類を決めてくれ」とソフトに要求されることも

データの種類



データの種類と要約値の種類

	意味	例	要約値	演算	
カテゴリ変数	順序尺度	大小関係にのみ意味がある 順序にも意味がある	薬効	度数 最頻値 中央値	大小の比較
	名義尺度	区別することに意味がある	疾患名 血液型	度数 最頻値	演算不可
連続変数	間隔尺度	数値の差のみに意味がある	温度 日付	度数 最頻値 中央値 平均値	加法 減法
	比尺度	数値の差と比に意味がある	年齢 身長	度数 最頻値 中央値 平均値	加法 減法 乗法 除法

データの要約値

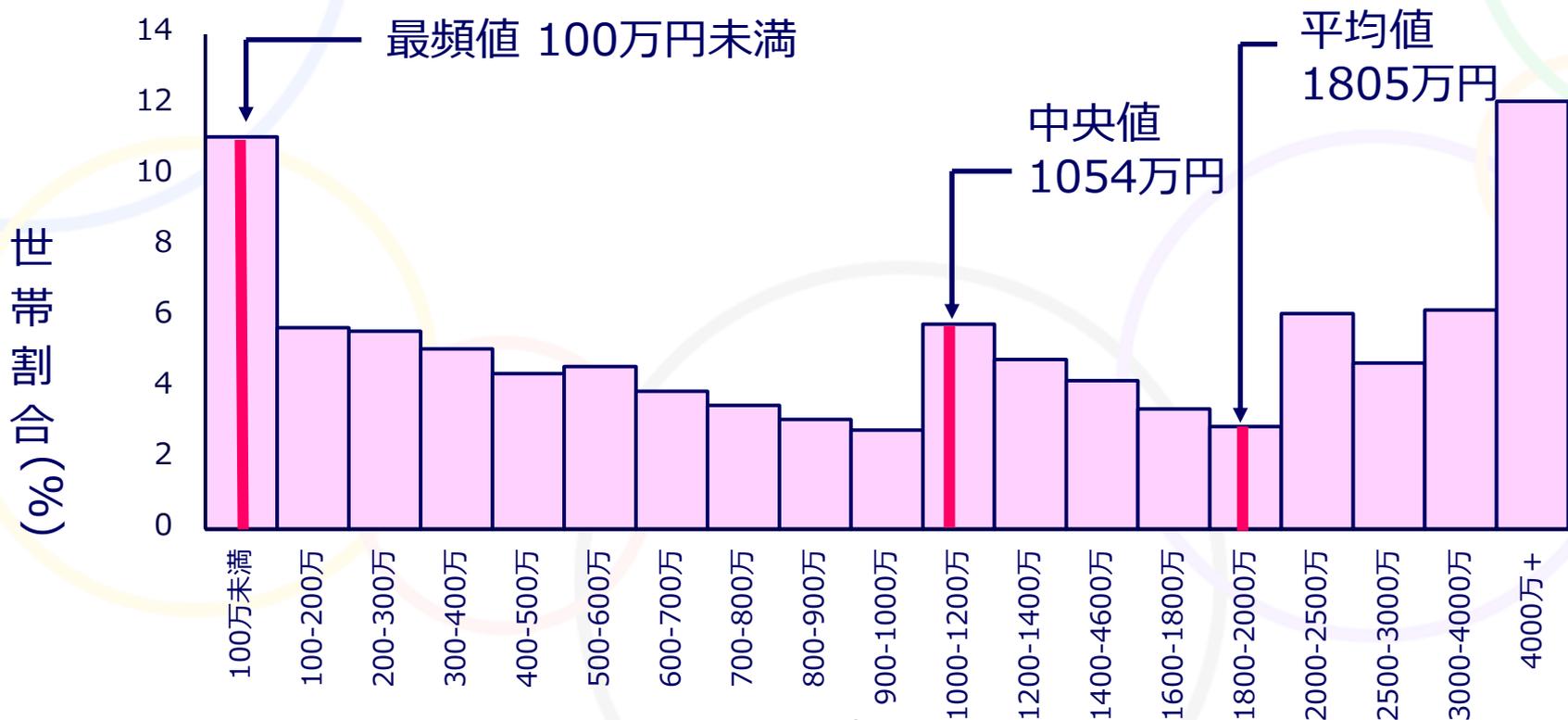
- **平均値 (Mean)** : 標本集団のデータの総和を標本数で割った値
- **中央値 (Median)** : 標本を大小順に並べて50%点 (=中央) にあたる値
- **最頻値 (Mode)** : 頻度が最も高い値

データの要約値

- どんな場合もとりあえず「平均」をとっておくとよい？



例) 貯蓄現在高階級別世帯分布 (二人以上の世帯)

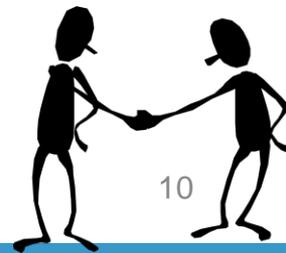


までー!



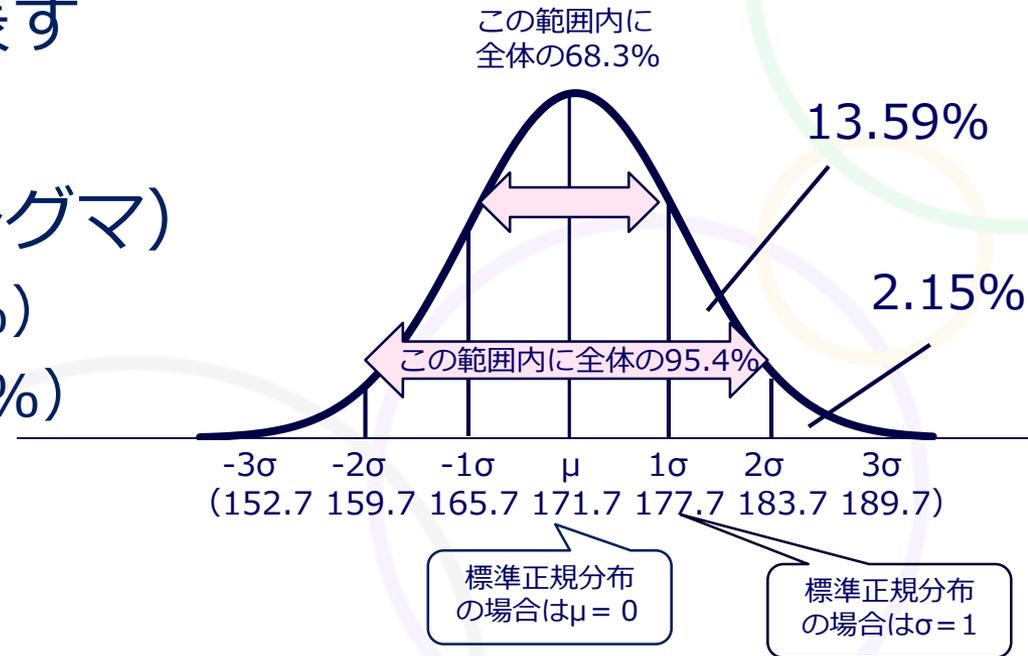
平均値が意味のある状況とは

- ✓ データが**正規分布**していること（対数処理などによる正規化を含む）
- ✓ **平均値 = 中央値 = 最頻値**
- ✓ 間隔尺度によるデータ
- ✓ データをプロットしてデータの分布を確かめる必要がある
- ✓ t検定や分散分析（ANOVA）は平均値に関する統計手法であり、平均値が意味のある状況にし
か使えない



正規分布とは

- 正規分布 $N(\mu, \sigma^2)$ と表す
- 平均値： μ （ミュー）
- 標準偏差（SD）： σ （シグマ）
 - $\pm 1SD$ に約 $2/3$ （68.3%）
 - $\pm 2SD$ に約 95%（95.4%）
 - $\pm 1.96SD$ に 95%
- 標準正規分布 $N(1, 0)$



例：平成27年国民健康・栄養調査より26-29歳の男性の身長は平均171.7cm、標準偏差（ σ ）6.0cmである。この度数分布が正規分布に従うとすると図のように示される。

それでは例題です

対象者一覧 (某グループ)

個人番号	生年月日	年齢	出身	活動開始年月日	卒業の有無	卒業年月日
1	1996年9月16日	20	山梨県	2014年1月1日	卒業	2016年9月30日
2	1991年7月22日	25	山梨県	2014年1月1日	在籍	
3	1993年10月6日	23	大阪府	2014年1月1日	在籍	
4	1996年10月8日	20	岩手県	2014年1月1日	卒業	2015年9月30日
5	1988年6月20日	28	山梨県	2014年1月1日	在籍	
6	1993年7月22日	23	奈良県	2014年1月1日	卒業	2016年9月30日
7	1994年11月3日	22	静岡県	2014年1月1日	卒業	2015年3月30日
8	1992年10月7日	24	北海道	2014年1月1日	在籍	
9	1991年12月7日	25	島根県	2014年1月1日	卒業	2016年3月30日
10	1998年10月4日	18	新潟県	2014年1月1日	在籍	
11	1989年11月25日	27	福井県	2014年1月1日	在籍	
12	1996年10月3日	20	愛媛県	2014年1月1日	在籍	
13	1993年7月22日	23	山形県	2014年1月1日	在籍	
14	1988年1月16日	29	秋田県	2014年1月1日	在籍	
15	2000年1月6日	17	秋田県	2014年1月1日	卒業	2014年6月30日

データの種類は何ですか

データ	データの種類	代表値
生年月日		
生月		
年齢		
出身		
(自分の) 好き嫌い		

データの種類は何ですか

データ	データの種類	代表値
生年月日	間隔尺度	度数、最頻値、中央値、平均値
生月	順序尺度	度数、最頻値、中央値
年齢	比尺度	度数、最頻値、中央値、平均値
出身	名義尺度	度数、最頻値
(自分の) 好き嫌い	順序尺度	度数、最頻値、中央値

データを表にしてみよう！

年齢階級	度数	都道府県	度数	生月	度数
17		山梨県		1	
18		大阪府		2	
19		岩手県		3	
20		奈良県		4	
21		静岡県		5	
22		北海道		6	
23		島根県		7	
24		新潟県		8	
25		福井県		9	
26		愛媛県		10	
27		山形県		11	
28		秋田県		12	
29					

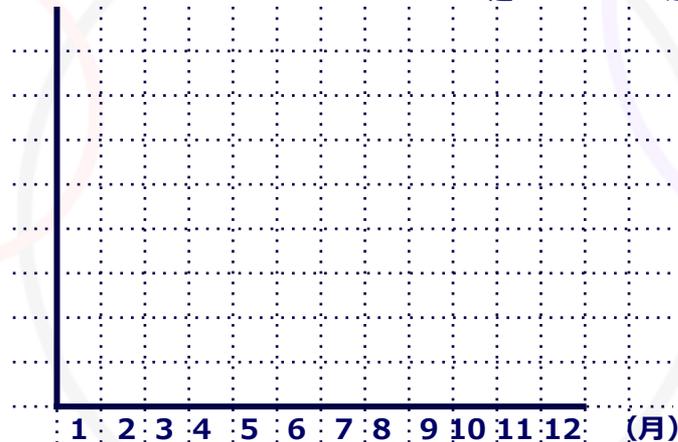
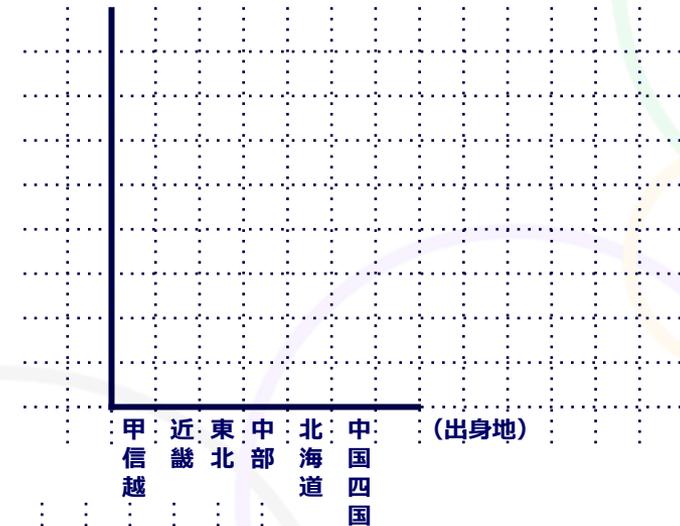
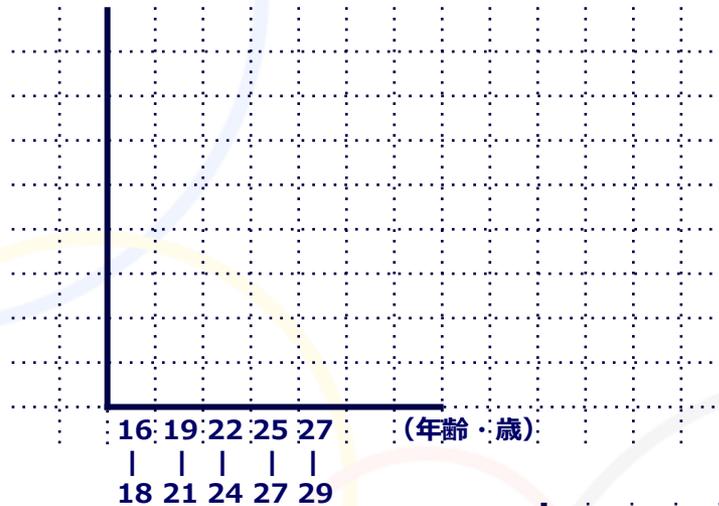
データを表にしてみよう！

年齢階級	度数	都道府県	度数	生月	度数
17	1	山梨県	3	1	2
18	1	大阪府	1	2	
19		岩手県	1	3	
20	3	奈良県	1	4	
21		静岡県	1	5	
22	1	北海道	1	6	1
23	3	島根県	1	7	3
24	1	新潟県	1	8	
25	2	福井県	1	9	1
26		愛媛県	1	10	5
27	1	山形県	1	11	2
28	1	秋田県	2	12	1
29	1				

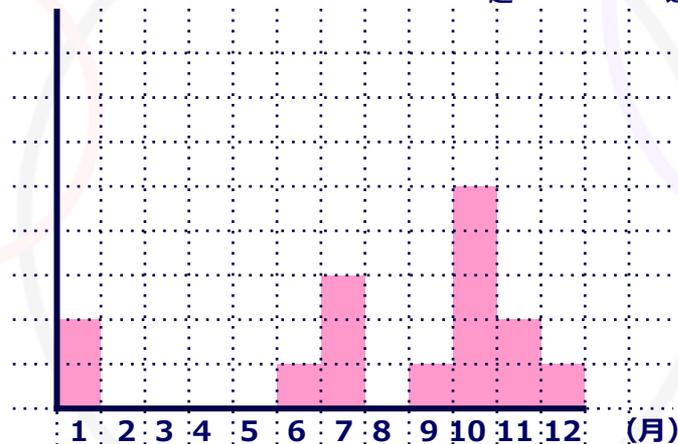
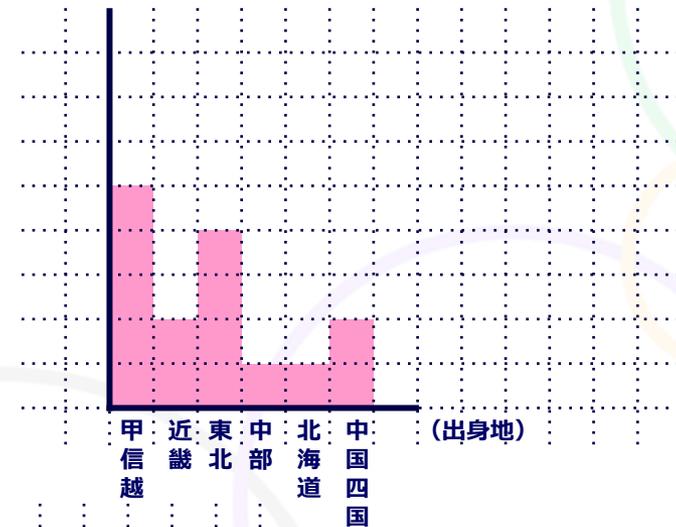
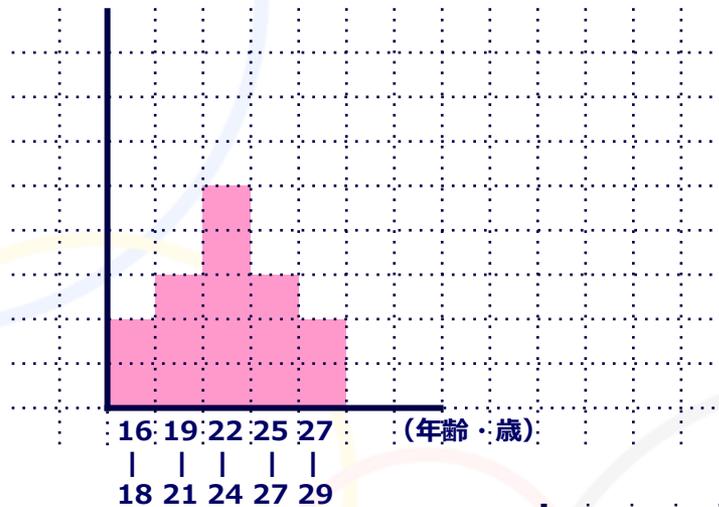
度数分布をグラフにしよう！



度数分布をグラフにしよう！



度数分布をグラフにしよう！



年齢の要約値を計算してみよう！

平均値	
中央値	
最頻値	
最小値	
最大値	

年齢の要約値を計算してみよう！

平均値	22.9
中央値	23.0
最頻値	20.0と23.0
最小値	17.0
最大値	29.0

割合、比、率

- **割合 (proportion)**

- 全体の中で特定の特徴を持つものが占める部分の大きさ

- **比 (ratio)**

- 異なる者同士を割り算で比較したもの

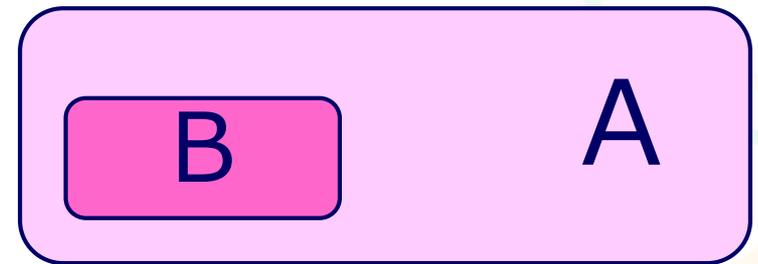
- **率 (rate)**

- 比の特殊な形で分母が時間になったもの
- 事象が生じる速さを示す指標

割合、比、率

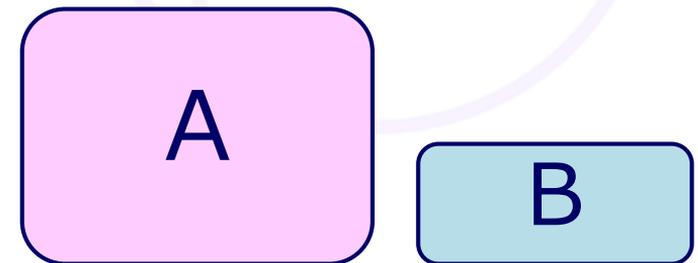
- **割合 (proportion)**

- BはAの中に含まれる
- B/A



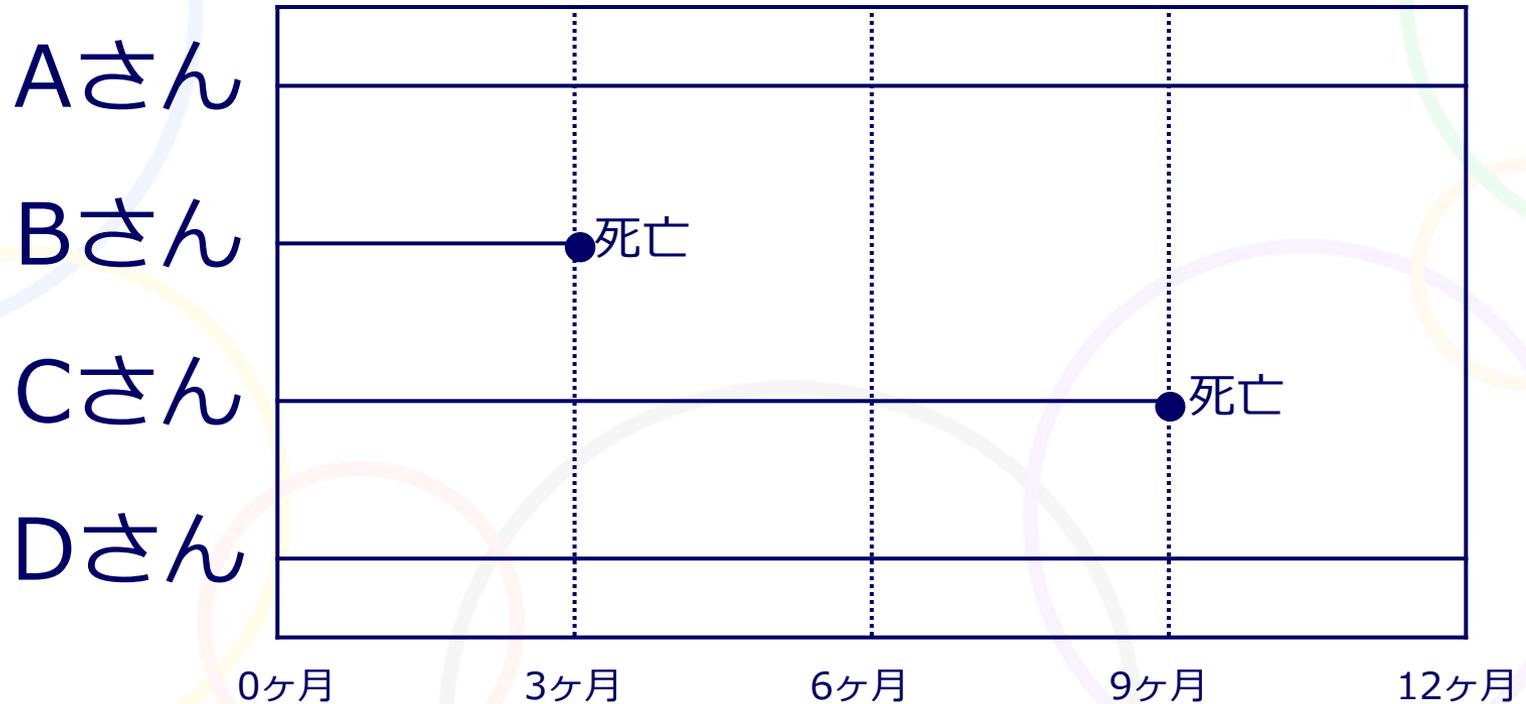
- **比 (ratio)**

- AとBは別物
- A対B



割合と率

(率は時間の概念を考慮に入れる)



- 死亡割合：2人／4人 = 0.5
- 死亡率：2人／(1 + 0.25 + 0.75 + 1) 人年 = 3人年あたり2人
- **分母の単位は何か？**

割合、比、率

- 日本語としての使用方法は曖昧（英語も）
- 用語として決まっているので、それぞれ定義を
しっかり覚えよう
 - 打率：安打数／全打数（割合）
 - 円周率：円の周長の直径に対する比（比）
 - 有病率：何人中何人が病気にかかっているか（割合）
 - 罹患率一定期間に新規に罹患した数を、その集団の人口
およびその期間の長さで割る（率）
 - 死亡率：一定期間に新規に死亡した数を、その集団の人口
およびその期間の長さで割る（率）

それでは例題です

全体に対する割合を出そう！

都道府県	度数	割合
山梨県	3	
大阪府	1	
岩手県	1	
奈良県	1	
静岡県	1	
北海道	1	
島根県	1	
新潟県	1	
福井県	1	
愛媛県	1	
山形県	1	
秋田県	2	
合計	15	

全体に対する割合を出そう！

都道府県	度数	割合
山梨県	3	$3/15 \times 100 = 20.00\%$
大阪府	1	6.66%
岩手県	1	6.66%
奈良県	1	6.66%
静岡県	1	6.66%
北海道	1	6.66%
島根県	1	6.66%
新潟県	1	6.66%
福井県	1	6.66%
愛媛県	1	6.66%
山形県	1	6.66%
秋田県	2	13.33%
合計	15	100.00%

比を出そう！

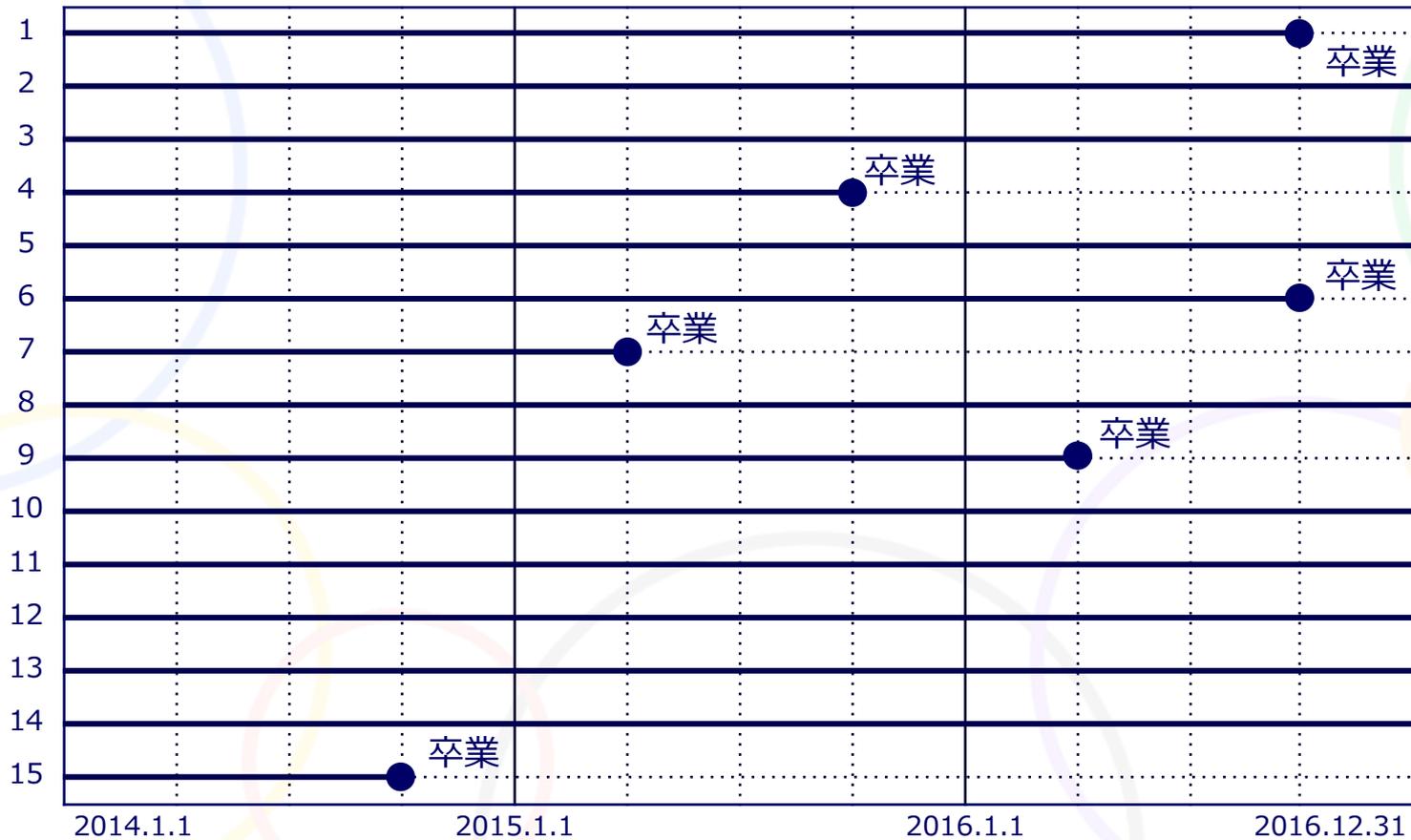
生月	度数		比
1	2	1月 : 6月	
6	1	1月 : 11月	
7	3	夏 : 秋	
9	1		
10	5		
11	2		
12	1		

比を出そう！

月	度数		比
1	2	1月 : 6月	2 : 1
6	1	1月 : 11月	2 : 2 = 1 : 1
7	3	夏 : 秋	4 : 6 = 2 : 3
9	1		
10	5		
11	2		
12	1		

卒業割合・卒業率をだそう！

個人番号	活動開始年月日	卒業の有無	卒業年月日
1	2014年1月1日	卒業	2016年9月30日
2	2014年1月1日	在籍	
3	2014年1月1日	在籍	
4	2014年1月1日	卒業	2015年9月30日
5	2014年1月1日	在籍	
6	2014年1月1日	卒業	2016年9月30日
7	2014年1月1日	卒業	2015年3月30日
8	2014年1月1日	在籍	
9	2014年1月1日	卒業	2016年3月30日
10	2014年1月1日	在籍	
11	2014年1月1日	在籍	
12	2014年1月1日	在籍	
13	2014年1月1日	在籍	
14	2014年1月1日	在籍	
15	2014年1月1日	卒業	2014年6月30日



- 2016年12月31日時点の
 - 卒業割合 : 6人/15人 = 0.4 (40%)
 - 卒業率 : 6人/38.5人年