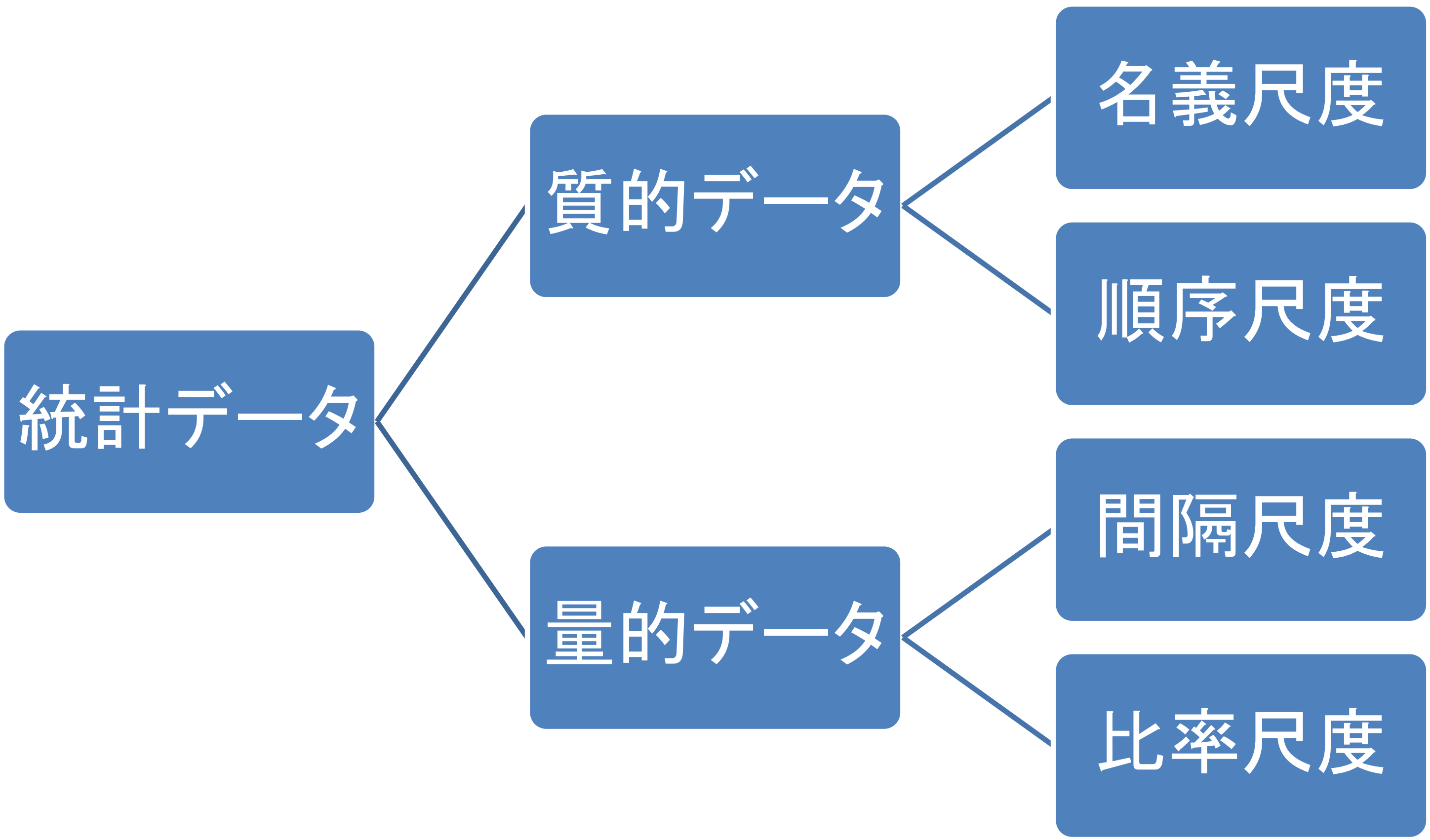
pp.18

第2章　統計的記述の基礎

統計データの分類

- 質的データ：　名義尺度・順序尺度
- 量的データ：　間隔尺度・比率尺度



平均

- 平均　　　　：統計データの総和を総数で割った値

平均値の定義式

$$\bar{x} = \frac{1}{n}\sum x_i$$

言葉に書き直すと…

$$平均 = \frac{1}{データ総数} * データの総和$$

$$平均 = \frac{データ全部の合計}{データの個数}$$

分散

- 分散　　　　：偏差平方の平均値

- 分散　(variance)

$$S_x^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$$

平均からの偏差

データ総数で割る

偏差平方の和

標準偏差

- 標準偏差：　　　分散の正の平方根

- 標準偏差　(standard deviation)

$$S_x = \sqrt{S_x^2} = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2}$$

変動係数

- 変動係数　：分布の拡がりを表す統計量

- **変動係数**（Coefficient of variation）
  - 標準偏差を平均で割る

$$C.V. = \frac{S_x}{\bar{x}}$$

  - 正の値をとる比率尺度のデータは
    平均値が大きいほど標準偏差も大きくなるため

標準化

- 標準化　　：統計データ内の相対的な位置の把握

- 標準化のしかた

$$z = \frac{x - \bar{x}}{S_x}$$

$$\text{標準化した変数} = \frac{\text{元の変数の平均からの偏差}}{\text{元の変数の標準偏差}}$$

- 標準化した変数(z)は
  - 平均＝0
  - 標準偏差＝1

共分散

- 共分散（covariance）
  平均からの偏差の積$(x_i - \bar{x})(y_i - \bar{y})$の和を
  データの総数nで割ったもの

- 共分散

$$S_{xy} = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$$

  – 統計データ全体で2変数の関連性を表す統計量
  – 2つの変数が**ともに量的データの場合のみ**計算可能

相関係数

- 相関係数（correlation coefficient）
  標準化した2変数の共分散
  単位をもたない

- 相関係数

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

共分散

xの標準偏差

yの標準偏差

  – 標準化する前の共分散を
  （xの標準偏差）×（yの標準偏差）で割る