

散布度 (dispersion)

- 代表値のほかに、重要な特性値として「**散布度**」がある。
- 平均値に対して、**どれくらいデータが散らばっているか**を示す。
 - 分布の裾の広がり具合
 - 平均値への集中の度合い

標準偏差

偏差 (deviation)

- 偏差 D は、各データと平均との差である。
 - + の偏差と - の偏差があるため、すべての偏差の合計は0になる。

$$D_i = \bar{x} - x_i$$

分散 (variance) と標準偏差 (standard deviation)

分析対象となる全体 (母集団) の分布のバラツキの度合いを求める場合には、代表的な散布度である、分散と標準偏差を用いる。

- 分散 s^2 (または σ^2) は、**偏差平方和** (偏差の二乗の和) をとって、その平均を求めたものである。
 - 全データの平均からのバラツキの程度を示す。

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n D_i^2 \end{aligned}$$

- 標準偏差 s は、分散の平方根を求めたものである。
 - 全データの平均からのバラツキの程度を示す (単位はデータと同じ)。

$$s = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n}}$$

- 標準偏差や分散の値が大きい場合はデータのバラツキが大きく、小さい場合はバラツキが小さい (データが同じ程度に揃ってる)

不偏分散 (unbiased variance) と不偏標準偏差 (unbiased standard deviation)

分析対象となる全体 (母集団) ではなく、対象の一部 (標本) の分布のバラツきの度合いを求める場合には、不偏分散と不偏標準偏差を用いる。

- 不偏分散 U^2 は、偏差平方和 (偏差の二乗の和) をとって、その平均を求めたものである。
 - 分散との違いは、分母は「標本数-1」であること。
 - データ全体についての平均値からのバラツきの程度を示す。

$$U^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2$$

- 不偏標準偏差 U は、分散の平方根を求めたもの
 - 全データの平均からのバラツきの程度を示す (単位はデータと同じ) 。

$$U = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n-1}}$$

標準偏差の和

n 組の資料 (データ) があるとき、資料全体の標準偏差は次のようになる。

$$s_T = \sqrt{\frac{\sum_{i=1}^n N_i (V_i + D_i^2)}{\sum_{i=1}^n N_i}}$$

- s_T ... 全体の標準偏差
- N_i ... 組目の資料の標本数
- V_i ... 組目の資料の分散
- D_i ... 組目の資料の偏差

範囲 (range)

- 範囲 R は、データの最大値 x_{max} と最小値 x_{min} との差で、データ全体の範囲を示す。
- ハズレ値の影響を受けやすい

$$R = x_{max} - x_{min}$$

四分位偏差 (quartile deviation)

- 四分位偏差はデータの変動の目安に利用される散布度で、代表値として中央値を用いたときに使われることがある。
- ハズレ値やデータ数に影響されにくい値である。

四分位偏差 = (第3四分位数 - 第1四分位数) / 2

平均偏差 (mean deviation)

- 平均偏差 M_{dev} は、偏差の絶対値を平均したもので、データと平均値とのずれの程度を示す。

$$\begin{aligned}M_{dev} &= \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \\ &= \frac{1}{n} \sum_{i=1}^n |D_i|\end{aligned}$$

変異係数 (coefficient of variance)

- 変異係数(変動係数) Cv は、標準偏差を平均で割ったもので、平均値に対する標準偏差の割合を示す(%表示)。
- 変異係数は相対的な散布度(割合を示す無名数で単位はない)で、平均値や標準偏差の異なる複数の種類のデータを比較するときに用いる。

$$Cv = s/\bar{x}$$

- 2つの系列(データの集まり)を比較するとき、次のような場合は相対的散布度が有利になる。
 - 双方の単位が同じで、平均がほぼ等しい
 - 双方の単位は同じだが、平均が違う
 - 双方の単位が違う