

◇ 教科書 p.44~p.56

●代表値

- ・ = データの値を集約して1つの数値として表したもの。

(例) 平均値, 中央値, 最頻値

平均値はすべてのデータを考慮に入れるが, 中央値はデータの真ん中の値のみが, 最頻値は最も多く現れる値が使われる。

- ・ = データの値のうち上位と下位をいくつか除外し, その上で平均をとったもの。
中央値もこの一つである。

●標準偏差

- ・ = データの散らばり具合を数値として表した指標。
データの散らばりが大きいほどこの値が大きく, 散らばりが小さいほどこの値は小さくなる。標準偏差が負の値をとることはない。

x_1, x_2, \dots, x_n を n 個のデータとすると,

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

$$s = \sqrt{s^2} \quad : \text{標準偏差}$$

●標準化

- ・ = データの平均を 0 に, 標準偏差を 1 に変換する操作。
これを行うことで, 平均値や標準偏差が異なる複数のデータ集合を対等に扱うことができる。標準化したデータの値は, 平均値より低い場合はマイナスに, 平均値より高い場合はプラスに, 平均値と同じ場合は 0 になる。

データ x_i を標準化した値を z_i とすると,

$$z_i = \frac{x_i - \bar{x}}{s}$$

偏差値はデータの平均値を 50 に, 標準偏差を 10 に変換した値。データ x_i の偏差値を w_i とすると,

$$w_i = 50 + 10 \times z_i = 50 + 10 \times \frac{x_i - \bar{x}}{s}$$

【課題】

エクセルの関数を用いて, 代表値, 標準偏差, 標準化の演習をしてみよう。

以下の関数は使えるようにしておこう。

- ・ 平均値…「average」, 中央値…「median」, 最頻値…「mode」, トリム平均…「trimmean」
- ・ 標準偏差…「stdev.p」