

代表値 (average)

- データの分布などの特徴を示す数値 (特性値) を「代表値」という。
- データ全体をひとつの値で代表させる値である。

平均値 (mean)

算術平均 (arithmetic mean)

- 算術平均 $\bar{x}$ は、データをすべて足しあわせ、標本数で割ったもの。
- 平均値でもっとも一般的。

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

幾何平均 (geometric mean)

- 幾何平均 Gm は、各変数の積に対して標本数のべき根を求めたもの。

$$Gm = \left(\prod_{i=1}^n x_i \right)^{1/n}$$

- 幾何平均の例
 - 5年間の物価上昇率が7%のとき、1年の平均上昇率は何%か？
 - 過去3年間の売上高の対前年比が120%、110%、130%のとき、平均の売上高の伸びは？

調和平均 (harmonic mean)

- 調和平均 Hm は、標本数を各変数の逆数の和で割ったもの。

$$Hm = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

- 調和平均の例
 - 山頂まで6kmの道のりを、往きは2km/hで、帰りは6km/hで歩いたとき、平均の速さはいくらか？
 - 車でドライブをして、最初の24kmは30km/h、次の24kmは40km/h、最後の24kmは60km/hで走った時、平均速度はいくらか？

中央値 (median)

中央値 (中位数)

- 中央値 Me は、データを大きさの順に並べたときに、中央にくる値のことである。
 - データ数が奇数のときは中央にくるデータの値になる。
 - データ数が偶数のときは中央にある2つのデータの平均の値になる。

$$Me = \begin{cases} x_m & \text{if } n \text{ odd, } m = (n+1)/2 \\ \frac{x_m + x_{m+1}}{2} & \text{if } n \text{ even, } m = n/2 \end{cases}$$

- 度数分布表がある場合は、階級や度数などから中央値を求めることもできる。

$$Me = l_m + \left(\frac{n}{2} - F \right) \frac{h}{f_m}$$

- n ... 標本数
- h ... 階級幅
- l_m ... m 番目の階級の下限值
- f_m ... m 組目の階級の度数
- F ... $m-1$ 番目までの累積度数

四分位数 (quartile)

- 四分位数は、ヒストグラムの面積を1/4ずつに分ける値である。
 - 中央値は、ヒストグラムの面積を半分に分ける値。
- 小さいほうから、第1、第2、第3四分位数という。
 - 中央値は、第2四分位数になる。

百分位数 (percentile)

- 百分位数 (パーセンタイル値) は、ヒストグラムの面積を1/100ずつに分ける値である。
 - 25パーセンタイル値は第1四分位数である。
 - 50パーセンタイル値は中央値 (で第2四分位数でもある) 。
- 度数分布表がある場合は、階級や度数などからパーセンタイル値 p を求めることもできる。

$$p = l_m + \left(\frac{n \times p}{100} - F \right) \frac{h}{f_m}$$

- n ... 標本数
- h ... 階級幅
- l_m ... m 番目の階級の下限值
- f_m ... m 組目の階級の度数
- F ... $m-1$ 番目までの累積度数

最頻値 (mode)

- 最頻値 Mo は、データのなかで**最も多く出てくる値**のことである。
 - 度数分布表がある場合は、もっとも度数の多い階級値を最頻値とする。
- 元のデータから最頻値を求める場合は、度数分布表から最頻値を求めることができる。

$$Mo = l_m + \frac{f_{m+1}}{f_{m-1} + f_m} \times h$$

- m ... 最大度数の階級
- h ... 階級幅
- l_m ... m 番目の階級の下限值
- f_m ... m 組目の階級の度数

代表値の特性

- 平均値はすべてのデータを反映している。
 - ハズレ値 (極端に飛び離れたデータ) があるとその影響を受けやすいため、ハズレ値の考慮が必要。

- 中央値 (四分位数や百分位数も) は分布上の位置 (中央など) を示す。
 - ハズレ値の影響を受けにくく、分布に偏りがある場合に優れている。
- 最頻値は、「データの多くはこのあたりにある」という説明をするのにわかりやすい。
 - ハズレ値の影響を受けにくい。