

単純集計 1変数:量的データ(1)

■ 要約統計量

- その変数全体または一部を数量的に要約する。

■ 代表値

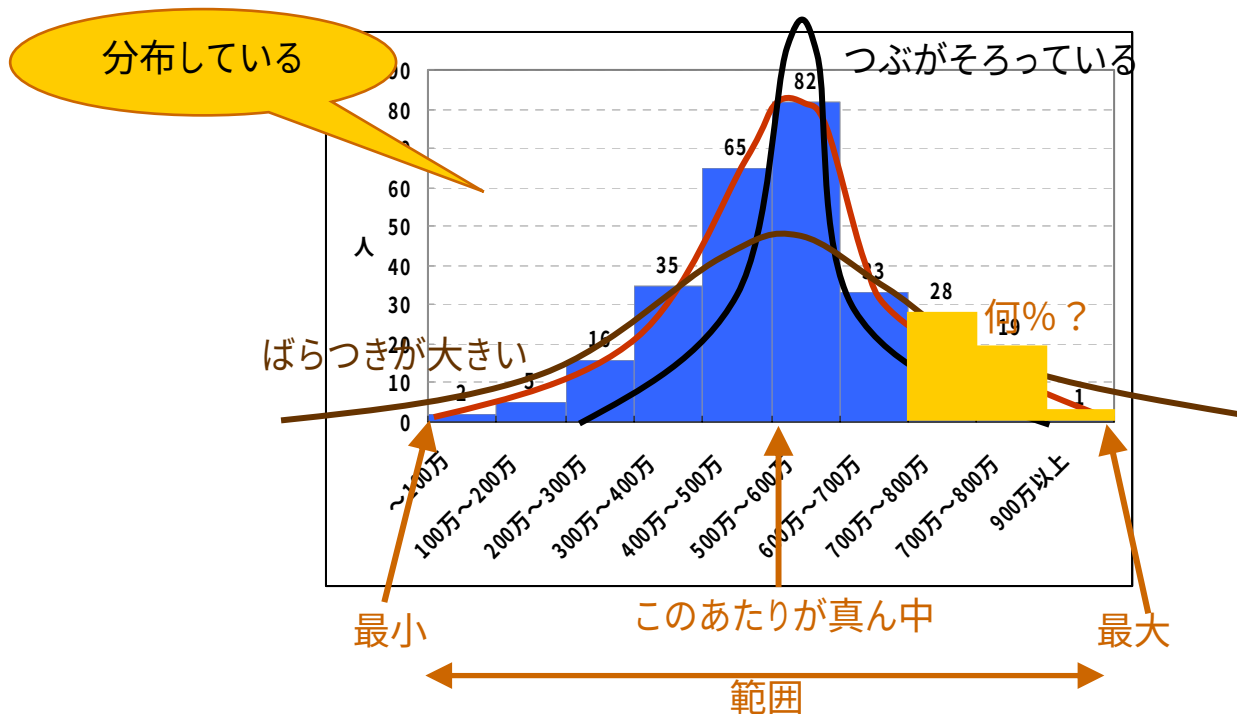
- 平均値
- 中央値
- 最頻値

■ 順序統計量

- 最大・最小
- 四分位
- パーセンタイル

■ 散布度

- 分散
- 標準偏差



単純集計 1変数:量的データ(2): 代表値

■ 算術平均

- データの総和をデータ数で割ったもの

$$A組 = (30 + 40 + 50 + 50 + 50 + 50 + 60 + 70) \div 8 = 50$$

$$B組 = (10 + 20 + 40 + 40 + 50 + 60 + 60 + 80 + 90) \div 9 = 50$$

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

A組 {30 40 50 50 50 50 60 70}

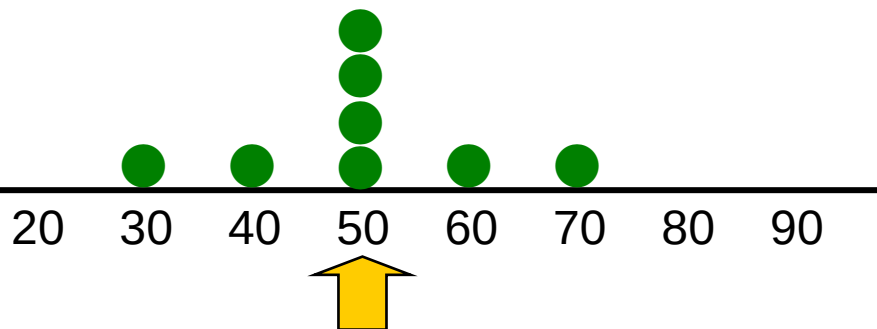
B組 {10 20 40 40 50 60 60 80 90}

■ 中央値

- データを大きさの順に並べたとき、ちょうど中央に位置する値

■ 最頻値

- 最も個数の多いデータ値



単純集計 1変数:量的データ(3):順序統計量

■ 最大・最小

- 文字通り, 最大値と最小値

A組 {30 40 50 50 50 50 60 70}



■ 四分位

- データの個数をちょうど四分の一ずつに分けるデータ値

- $1/4 \cdots$ 第1四分位数 $8 \times 0.25 = 2$ (個) $\rightarrow 40$
- $2/4 \cdots$ 第2四分位数 = 中央値 $8 \times 0.50 = 4$ (個) $\rightarrow 50$
- $3/4 \cdots$ 第3四分位数 $8 \times 0.75 = 6$ (個) $\rightarrow 50$

■ パーセンタイル

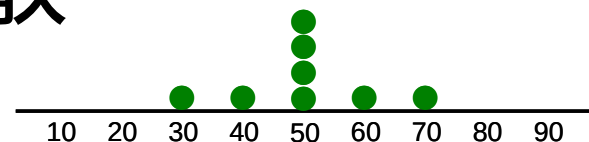
- そのデータ値よりも小さな値が全体の $\alpha\%$ あるとき, そのデータ値のことを α パーセンタイルと呼ぶ。

単純集計 1変数:量的データ(4):分散

A組 {30 40 50 50 50 50 60 70}

	↓	↓	↓	↓	↓	↓	↓	↓
偏差	-20	-10	0	0	0	0	10	20
	↓	↓	↓	↓	↓	↓	↓	↓
偏差平方	400	100	0	0	0	0	100	400

総和
平均 $\div 8$ 125

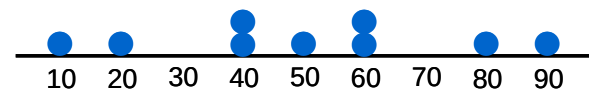


小 125

B組 {10 20 40 40 50 60 60 80 90}

	↓	↓	↓	↓	↓	↓	↓	↓	↓
偏差	-40	-30	-10	-10	0	10	10	30	40
	↓	↓	↓	↓	↓	↓	↓	↓	↓
偏差平方	1600	900	100	100	0	100	100	900	1600

総和
平均 $\div 9$ 600



大 600

分散

平均からのズレの2乗の平均

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

単純集計 1変数:量的データ(5):標準偏差

■ 標準偏差

□ 分散の平方根

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

