代表値
（教科書5章）

北九州市立大学経済学部

齋藤 朗宏

今日の内容

- 代表値
- 散らばりの指標
- 標準化と偏差値得点

- 実習

要約統計量

- グラフでは直感的な理解が可能だが，データを報告する場合には，より厳密に数値で示す必要もある．報告するにあたっては，最初に平均・分散といった要約統計量を示すのが一般的である．

平均値

- 一般的な「日本人成人女性の身長」とは？
- 一般的な「北九州市八幡西区，5月の最高気温」とは？
- 一般的な「大学1年終了時の取得単位数」とは？

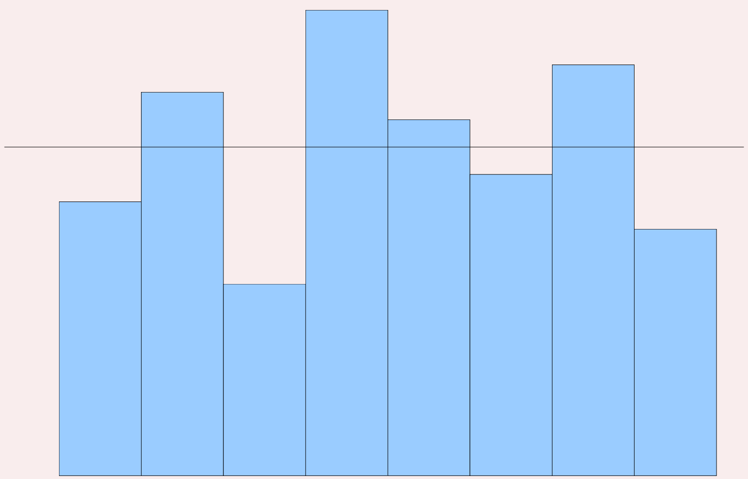
- 一般的，とは何か？
  - （教科書5.2）

平均値

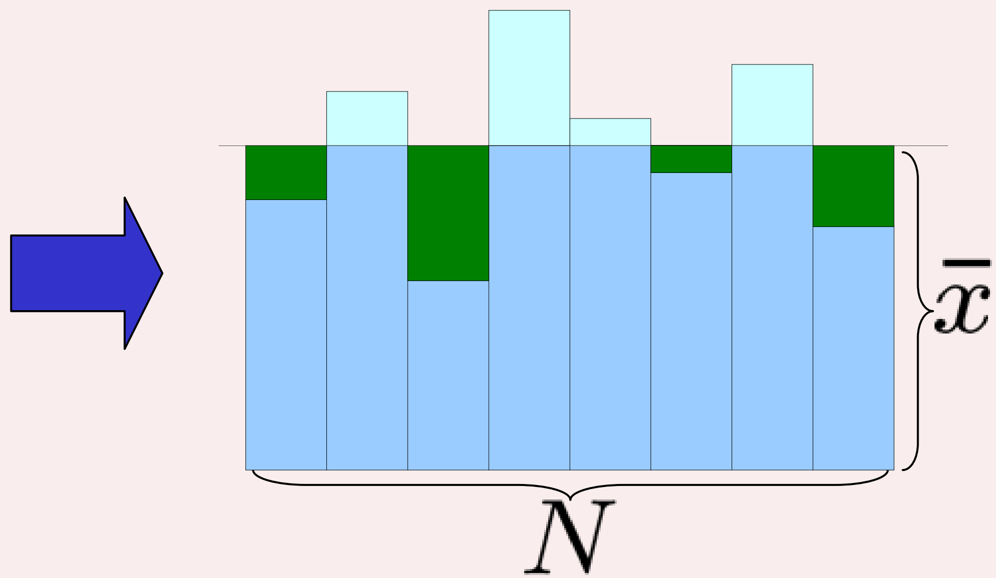
- 集団を代表している必要がある.
  - その集団の傾向を明確に示す値である.
- 平均値がよく用いられる.

平均値

- ある変数xに対するi 番目のオブザベーションの結果をx_iとする．また，オブザベーション数をNとする．
- 算術平均値（ある集団を代表する値）



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



ワリカンの考え方

平均値

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

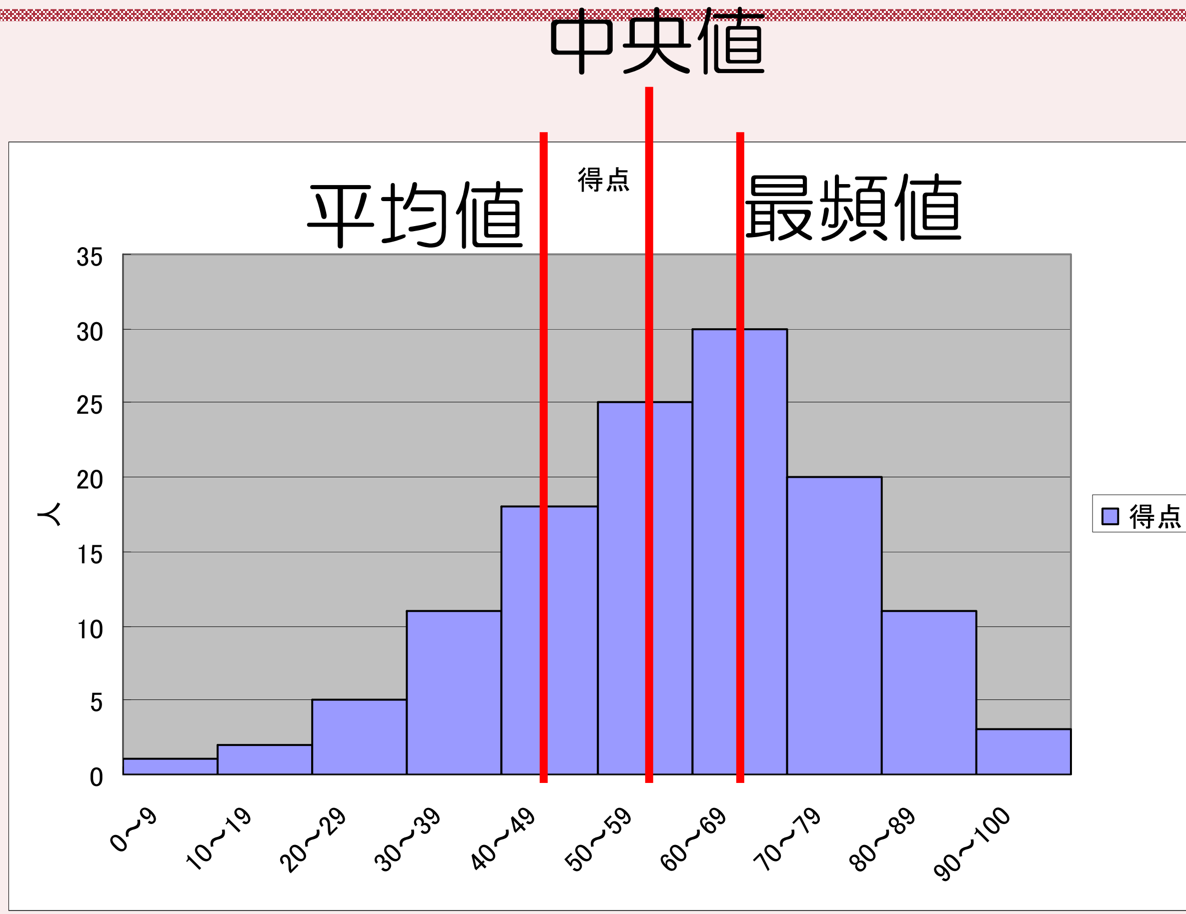
∑の中身をN個分足してNで割ることで，平均を算出するという意味を持つ.

この部分の平均を出す.

中央値・最頻値

- 算術平均は外れ値の影響が大きい.
- $x_1, x_2, \cdots, x_N$ について，値の小さなものから順に並べ替えたものを以下とする.
  $x_{(1)}, x_{(2)}, \cdots x_{(N)}$
- このとき, N が奇数であれば x_m を中央値と呼ぶ．ただし, $m = (N+1)/2$ である.
- N が偶数の場合には, $(x_m + x_{m+1})/2$ を中央値と呼ぶ．ただし, $m = N/2$ である.
- 度数分布表・ヒストグラムで度数が最大である階級のことを最頻値と呼ぶ.

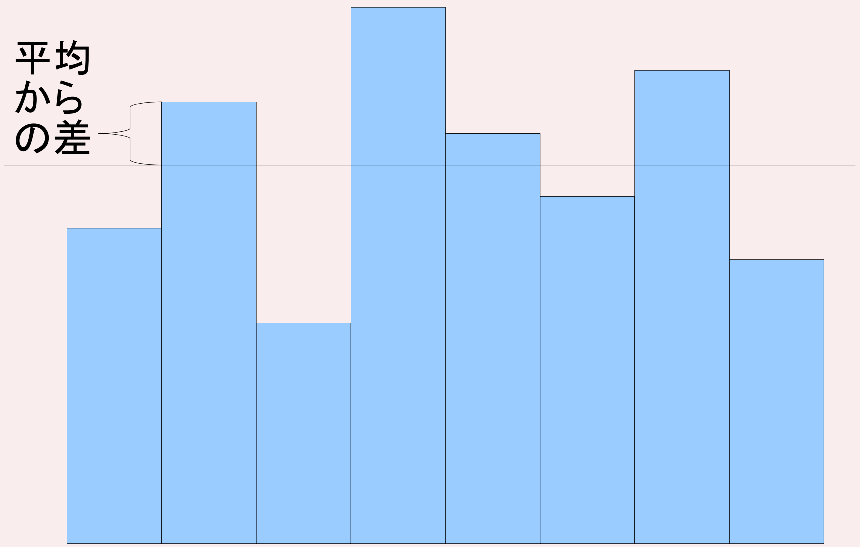
平均値・中央値・最頻値の関係

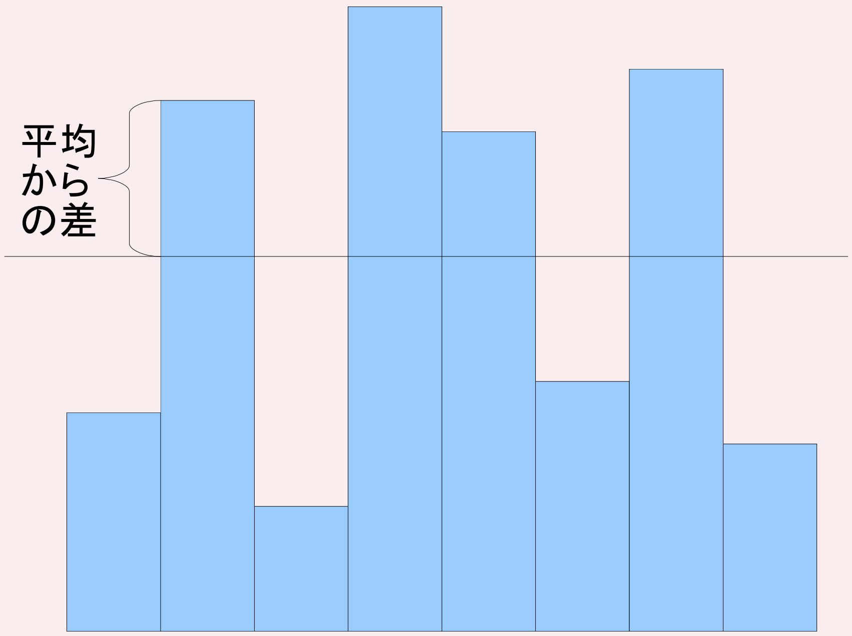


- 算術平均値は外れ値の影響を受けやすく，逆に最頻値は影響を全く受けない．

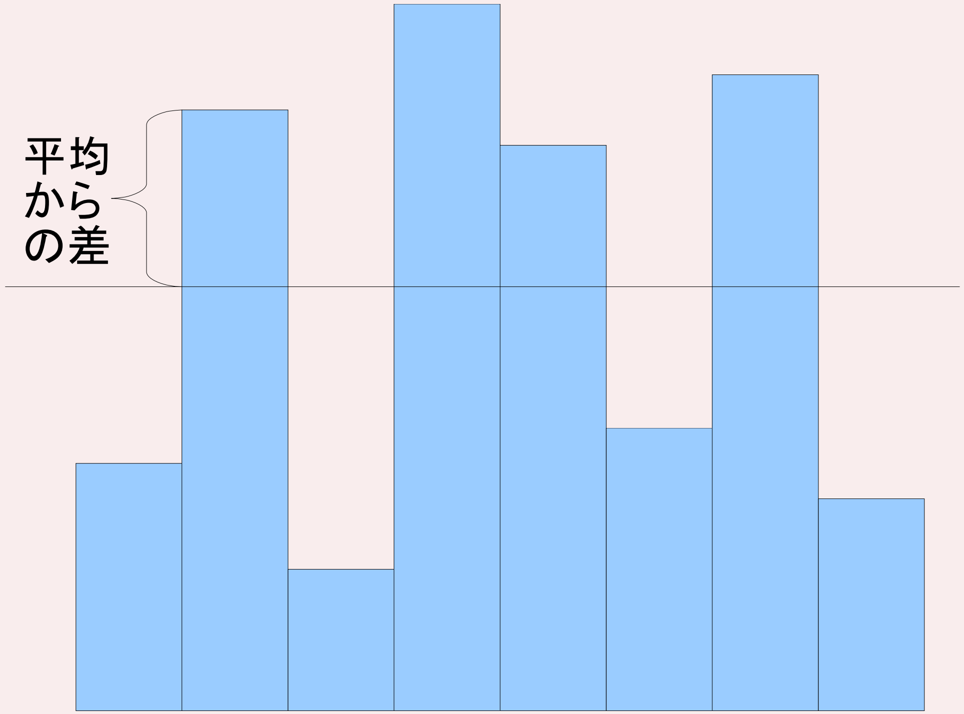
散らばりの指標

- 同じ平均身長170cmだったとしても，190cmと150cmの集団の方が165cmと175cmの集団よりも散らばりが大きい.





平均偏差



- “平均からの差” の平均を取るのが簡単.

$$d_x = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|$$

平均偏差

$$d_x = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|$$

Σの中身をN個分足してNで割ることで，平均を算出するという意味を持つ．

この部分の平均を出す．

つまり，$|x_i - \bar{x}|$ の平均値が平均偏差．

個々の値が平均値から平均的にどれだけ離れているのかがわかるので，散らばりの指標となる．

分散

- 絶対値は数学的には扱いが難しい．そこで，平均からの差を2乗したものの平均を分散とする．

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N} \underline{(x_i - \bar{x})^2}$$

$(x_i - \bar{x})^2$ の平均値が分散．

個々の値が平均値からどれだけ離れているのか，その大きさがわかるので，散らばりの指標となる．

２乗について

$$\begin{bmatrix} -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \end{bmatrix}$$

$|x| > |y|$ならば，$x^2 > y^2$であり，
$|x| = |y|$ならば$x^2 = y^2$が成立する．

問題

1. ４人の身長が190cm，170cm，170cm，150cmだったとする．このとき，この4人の身長の平均と分散は？
2. 180cm，170cm，170cm，160cmである4人の身長の平均と分散は？

正解（１）

$$
\begin{aligned}
\bar{x} &= \frac{190+170+170+150}{4} = 170 \\
\sigma^2 x &= \frac{(190-170)^2 + \cdots + (150-170)^2}{4} \\
&= 200
\end{aligned}
$$

正解（２）

$$
\begin{aligned}
\bar{y} &= \frac{180+170+170+160}{4} = 170 \\
\sigma^2 y &= \frac{(180-170)^2 + \cdots + (160-170)^2}{4} \\
&= 50
\end{aligned}
$$

標準偏差

- 分散は，2乗したものの平均であるため扱い辛い部分がある．そこで，平方根をかけて元に戻す．戻したものは標準偏差と呼ばれる．

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

$$\sigma_x = \sqrt{\sigma_x^2}$$

問題

1. ４人の身長が190cm，170cm，170cm，150cmだったとする．このとき，この4人の身長の平均と分散・標準偏差は？
2. 180cm，170cm，170cm，160cmである4人の身長の平均と分散・標準偏差は？

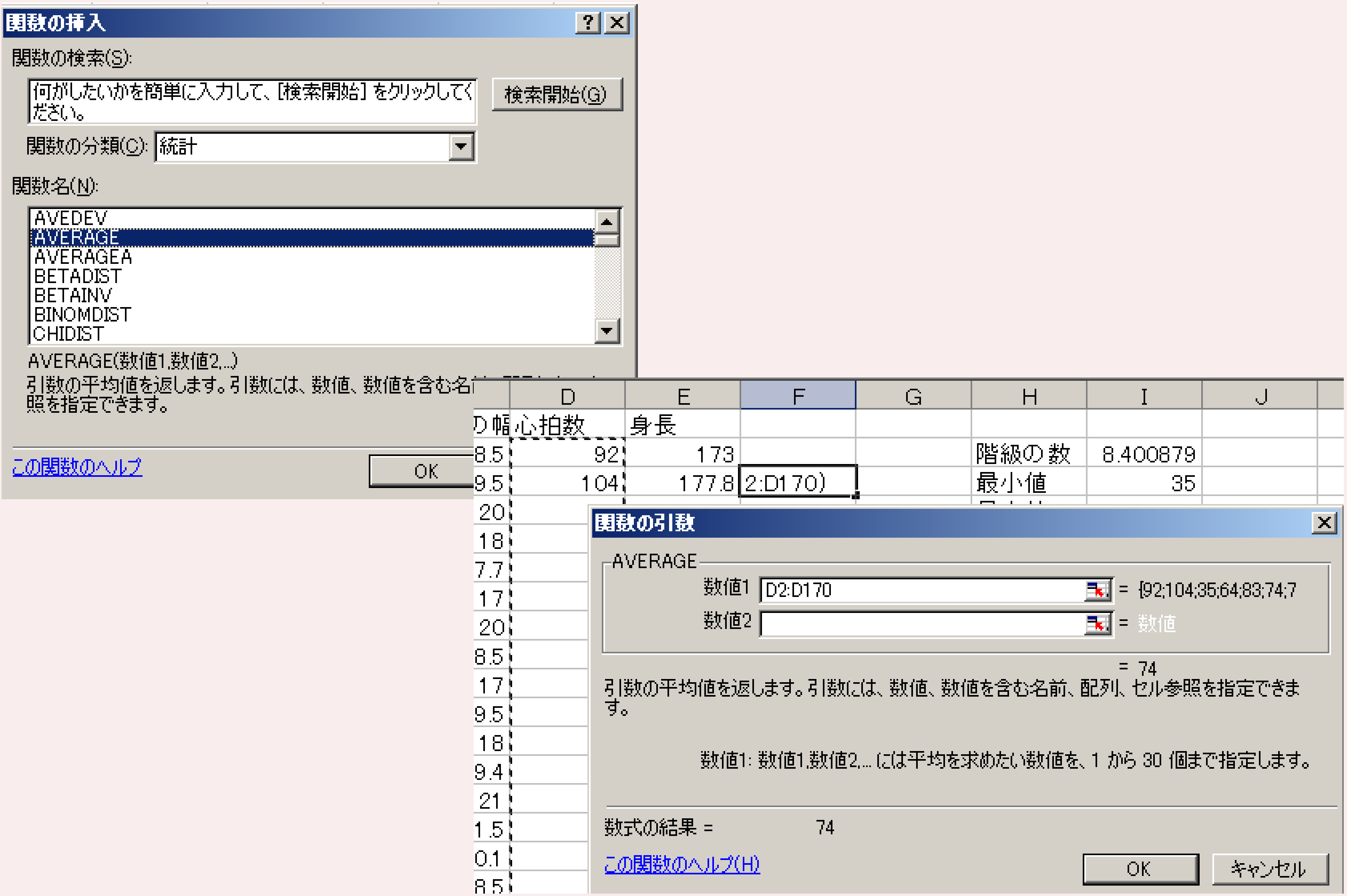
正解（１）

$$
\begin{aligned}
\bar{x} &= \frac{190+170+170+150}{4} = 170 \\
\sigma^2 x &= \frac{(190-170)^2+\cdots+(150-170)^2}{4} \\
&= 200 \\
\sigma x &= \sqrt{200} \simeq 14.14
\end{aligned}
$$

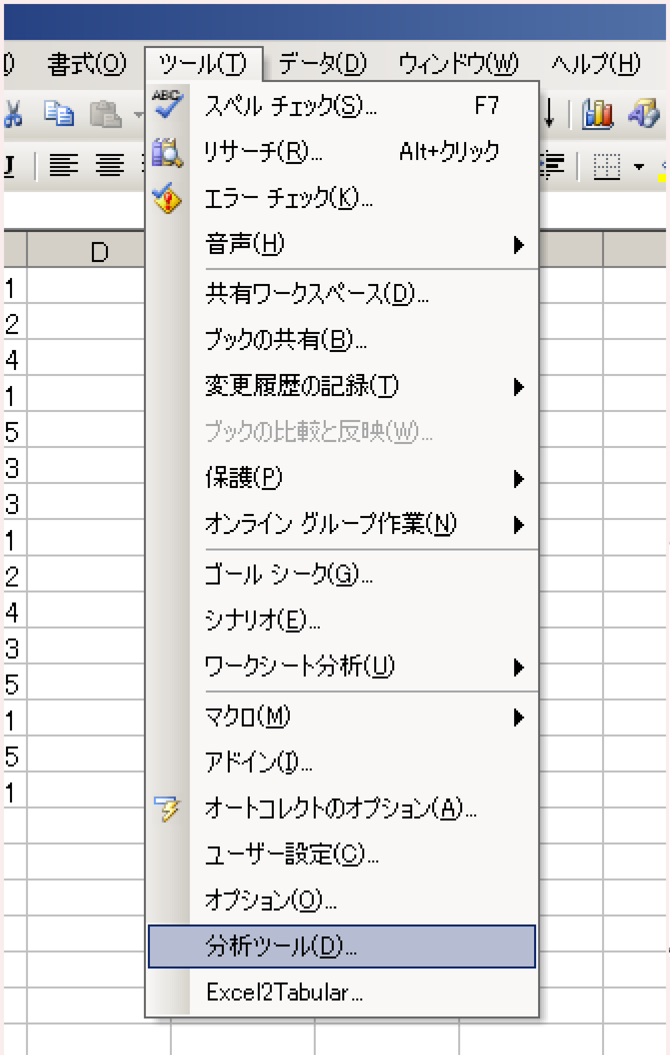
正解（２）

$$
\begin{aligned}
\bar{y} &= \frac{180+170+170+160}{4} = 170 \\
\sigma^2 y &= \frac{(180-170)^2 + \cdots + (160-170)^2}{4} \\
&= 50 \\
\sigma y &= \sqrt{50} \simeq 7.07
\end{aligned}
$$

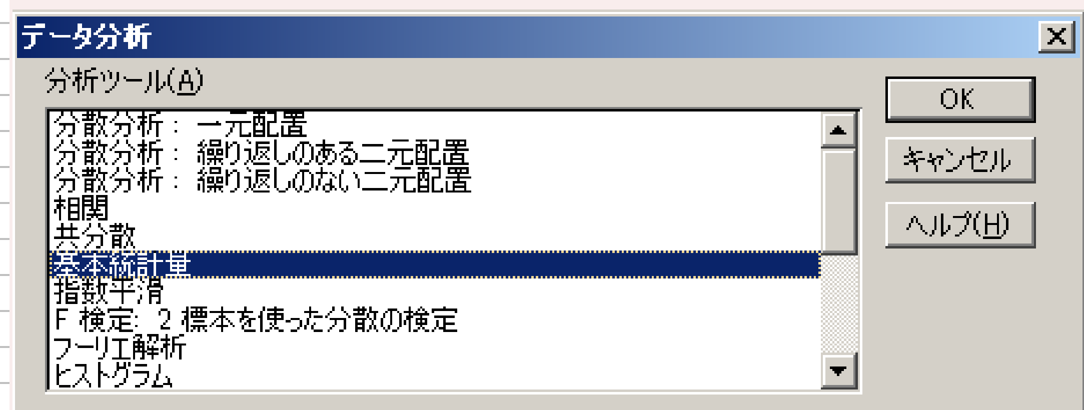
Excelによる平均，分散，標準偏差



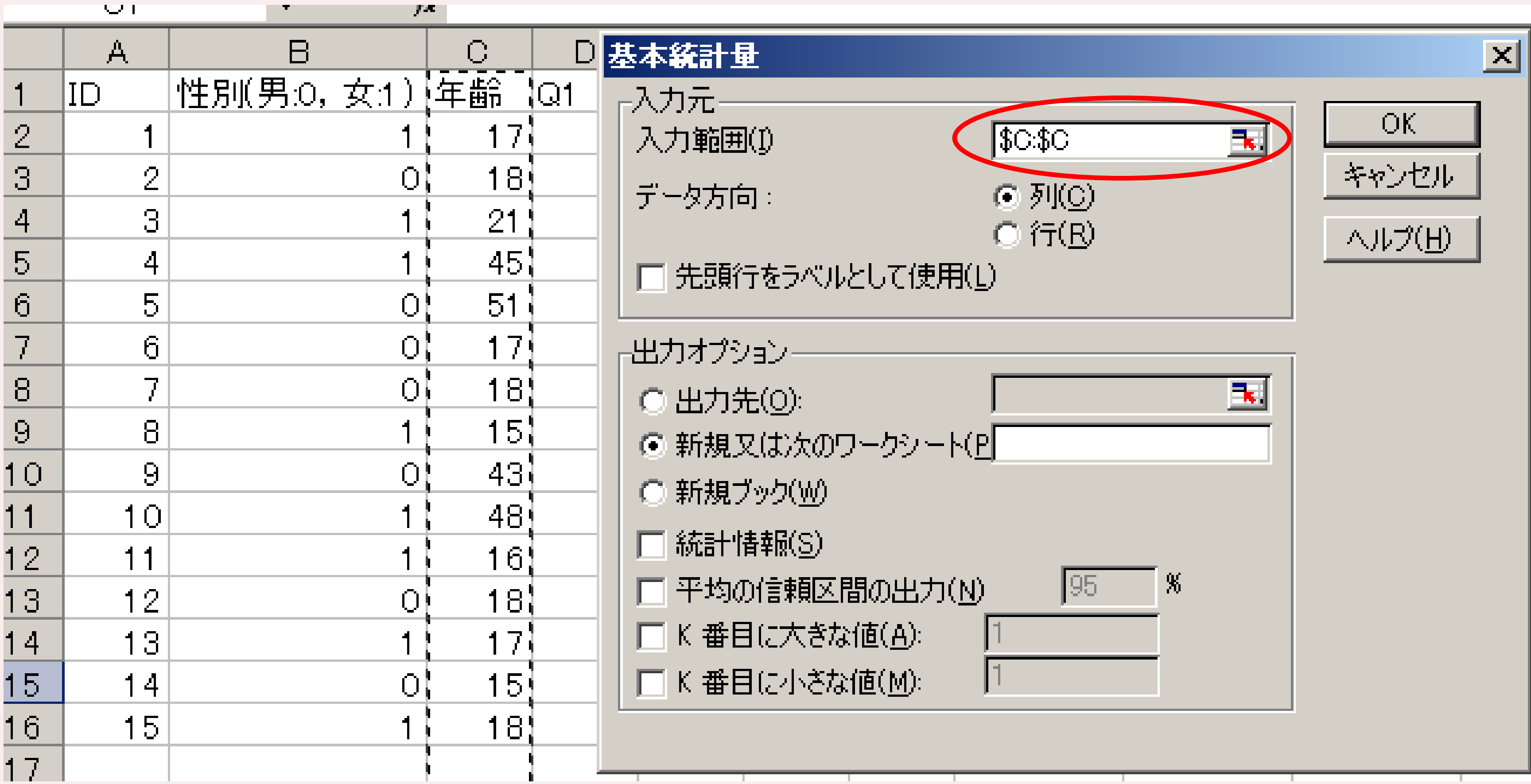
Excelによる要約統計量の算出



- ツール→分析ツールをクリック.
- 基本統計量を選択し，OKをクリック.

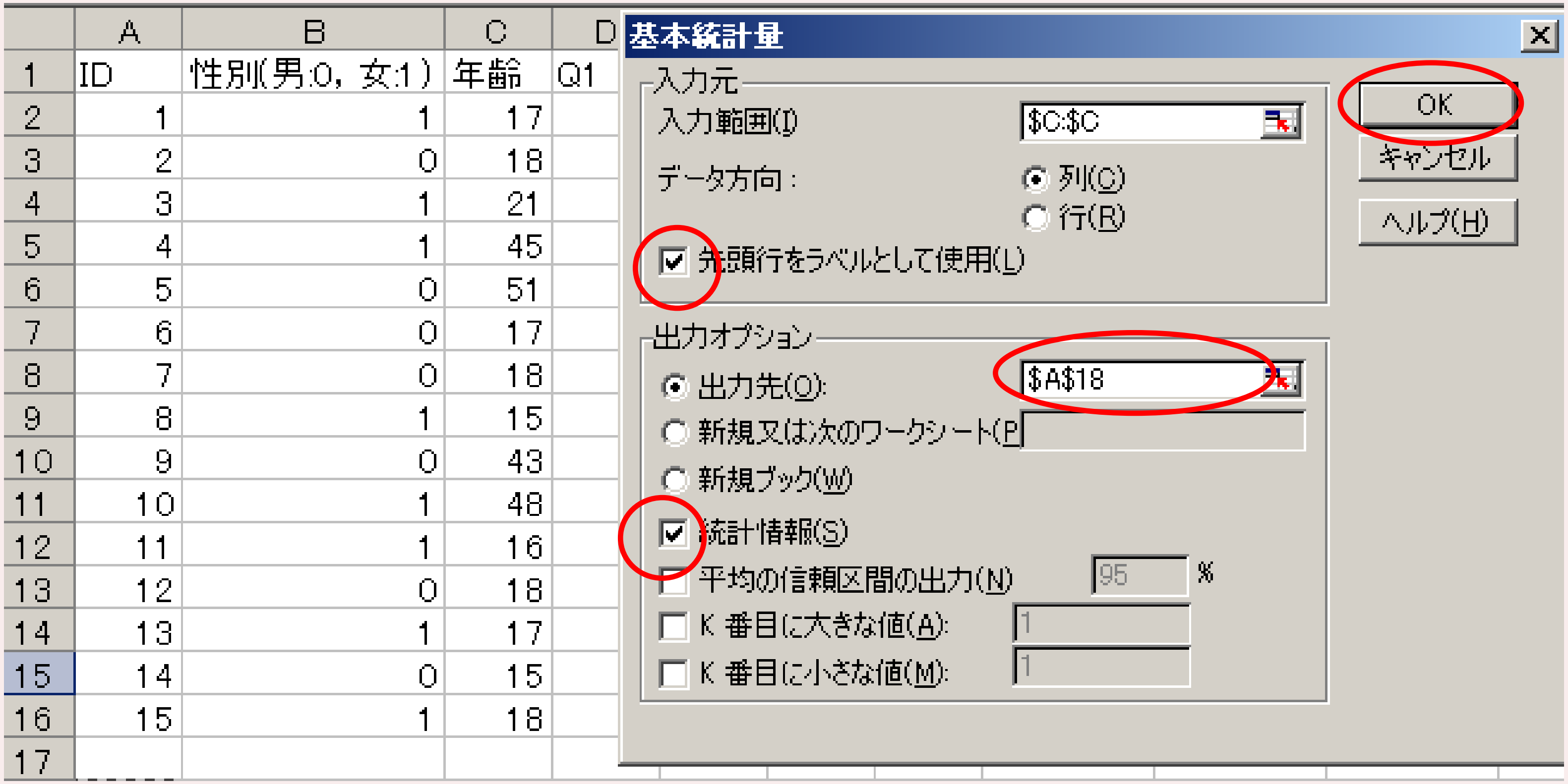


Excelによる要約統計量の算出



- 入力元の白い部分をクリックし，分析したい行（ここではC）を選択．

Excelによる要約統計量の算出



- 統計情報にチェック，先頭行が項目名の場合先頭行をラベルとして使用にチェックを入れ，必要に応じて出力先を変更しOK.

データの標準化

- データから平均を引いて標準偏差で割ることで，平均0，標準偏差1に調整することができる．これを標準化と呼ぶ．

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

- 標準化された値について，10倍して50を足すと偏差値得点となる（平均50，標準偏差10）．

$$T_i = 10z_i + 50$$

標準化と偏差値算出

ZE ▾ ✕ ✓ fx =STANDARDIZE(B2,F2,F6)

	B	C	D	E	F
	心拍数	標準化	偏差値		
1	92	=STANDARDIZE(B2,F2,F6)			74
2	104	STANDARDIZE(x, 平均, 標準偏差)			72
3	35			最頻値	80
4	64			分散	131.6331
5	83			標準偏差	11.47315
6	74			最大値	104
7	72			最小値	35
8	90				

心拍数	標準化	偏差値		
92	1.568881	=C2*10+50		74
104			中央値	72
35			最頻値	80
64			分散	131.6331
83			標準偏差	11.47315
74			最大値	104
72			最小値	35

実習

- chp05_a.xlsのお年玉データに関して，算術平均値，中央値，最頻値，分散，標準偏差，最大値，最小値を分析ツールを用いずに求めよ.
- 分析ツールで要約統計量を算出し，上の結果と比較せよ.
- 「お年玉」が10000円である場合の標準化得点，偏差値を算出せよ.