



総務省統計局

社会人のためのデータサイエンス演習

第2週：課題の補講 Excel分析演習

講師名：菅 由紀子

講義内容

補講

- 課題の補講 Excel分析演習

※Microsoft® , Microsoft® Office Excel® は、米国 Microsoft Corporation の、米国およびその他の国における登録商標または商標です。

※本資料は、Microsoft Corporation と提携しているものではなく、また、Microsoft Corporationが許諾、後援、その他の承認をするものではありません。

※本資料の本文では、©、®、™などの表記は割愛いたします。

※本資料ではデータ分析ツールとして Microsoft® Office Excel® 2019 を利用しています。

Microsoft® Office の他のバージョンや他の分析ツールを利用している場合は、ヘルプやインターネットなどで各自で調査し、該当機能に置き換えて参照してください。

第2週のまとめ

Analysisの具体的手法

分布 / 比較 / 傾向

KGIとそれに関連する要因の関係を分析する

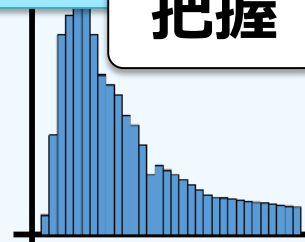
分布

基本統計量

- ・ 平均値
- ・ 標準偏差
- ・ 最頻値
等

把握

顧客数

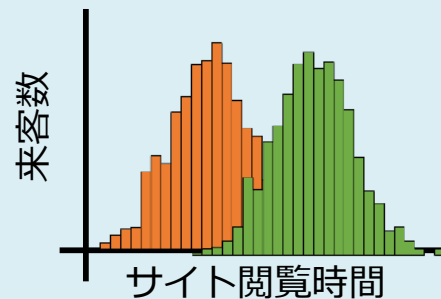


購入金額
ヒストグラム

比較

名義 × 比例

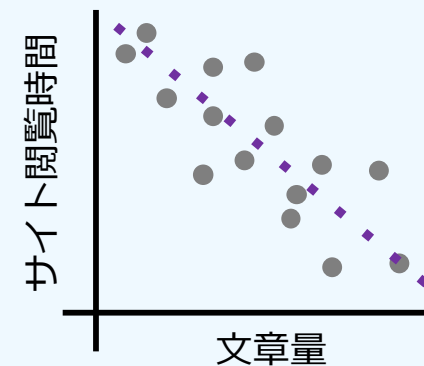
サイトA サイトB



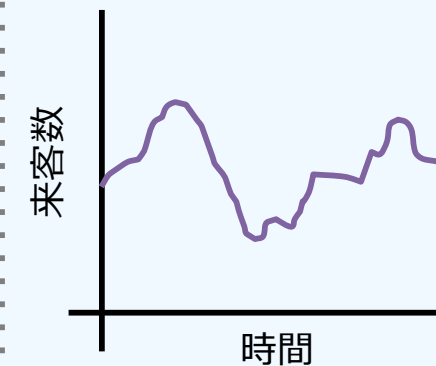
ヒストグラム

傾向

比例 × 比例



散布図



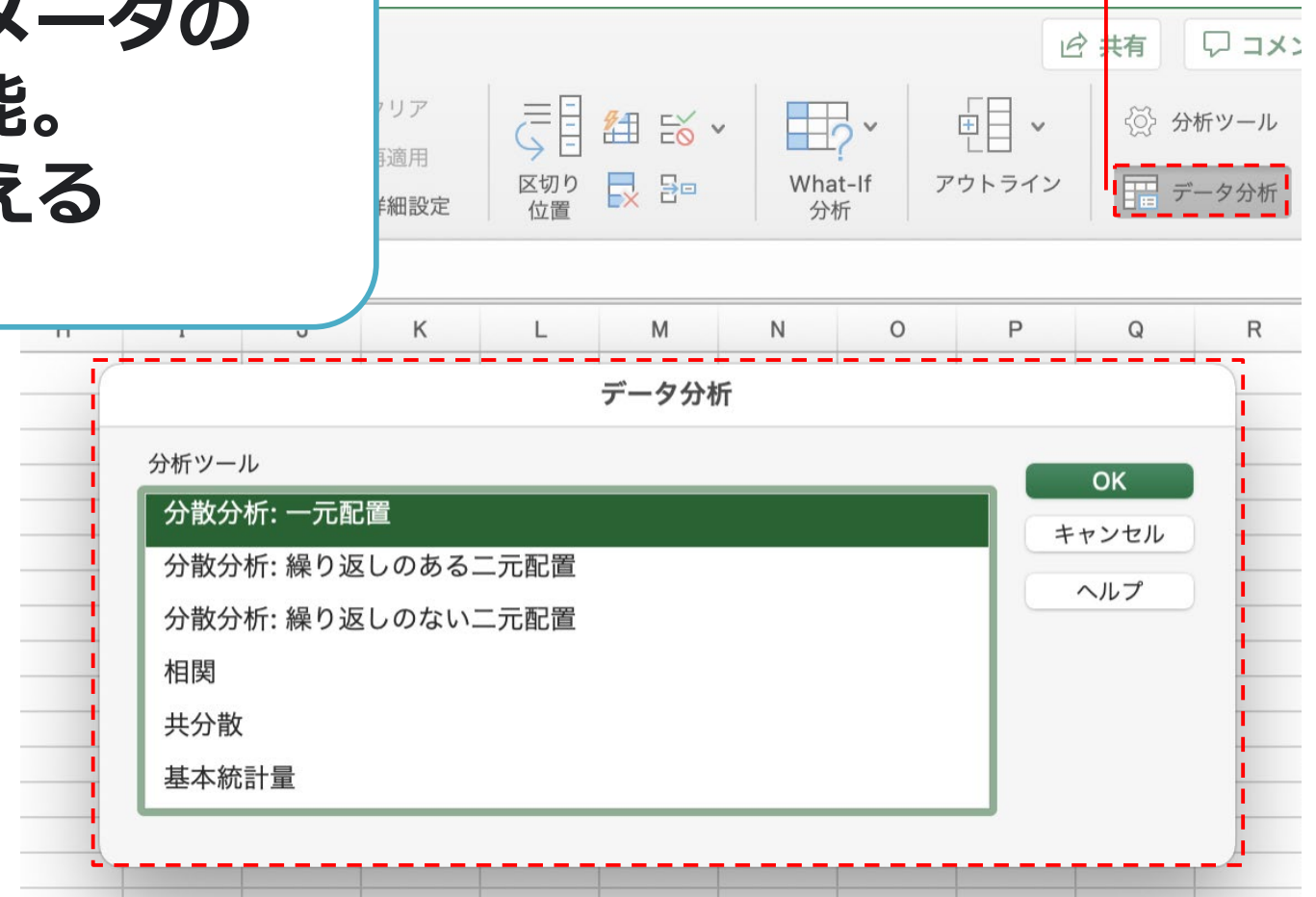
時系列

分析ツールの有効化

Excelの分析ツールとは

- ✓ エクセルのアドイン機能の一つ
- ✓ 関数などを使用せずに簡単なパラメータの入力のみでデータ分析の実行が可能。
- ✓ 初心者でも分析の実行が簡単に行える

分析ツールを有効化すると
データのメニュー内に
「データ分析」が表示される



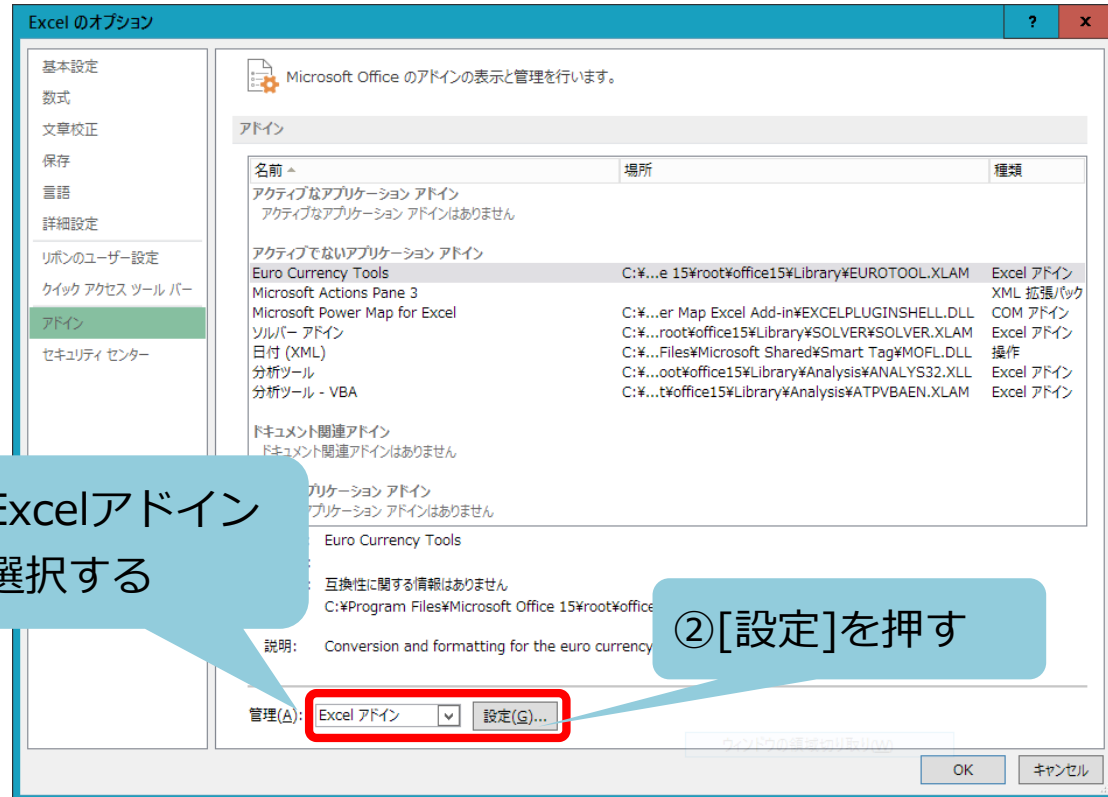
分析ツールの有効化

手順①

Excelのメニューから[ファイル]→[オプション]
→ [アドイン] と進み、
[管理]ボックスから[Excelアドイン]を選択して
[設定]ボタンを押す

手順②

[分析ツール]にチェックを入れて、[OK]を押すと
分析ツールアドインを有効化となる。



①Excelアドイン
を選択する

②[設定]を押す



③[分析ツール]に
チェックを入れる

④[設定]を押す



基本統計量の算出

基本統計量とは

データの基本的な特徴を記述したり要約するために必要な指標。
位置を表す値、ばらつきを表す値、分布の形状を表す値などさまざまな特徴を持つ。

代表値について（動画：2-3 1変数の状況の把握②（代表値の活用）より）

位置を示す代表値

- 平均値
- 中央値
- 最頻値

ばらつきを示す代表値

- 標準偏差

分布の形を示す代表値

- 尖度
- 歪度

平均値

分布の位置や重心の状況を表す値

中央値

分布をデータの全データサンプルを下位半分と上位半分に分ける値

最頻値

最も多い度数（頻度）を示す値

標準偏差

分布の平均値の周りのばらつき度合いを表す値

尖度

ピークへの集中度合いを表す指標。

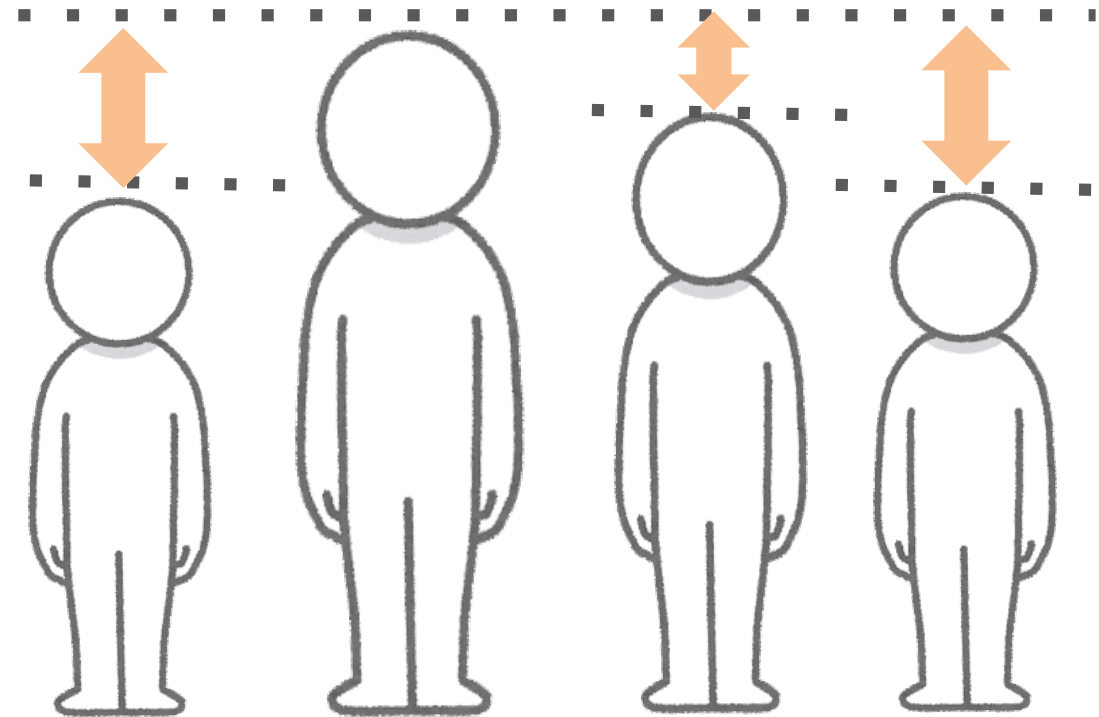
歪度

分布の左右への非対称性を表す代表値
0に近いほど左右対称の分布であることを示す。

基本統計量の算出

あるグループの男性の身長を測ったデータがあります。
データの特徴を把握するために、基本統計量を算出してみましょう。

身長 (cm)
183
167
175
164
174
170
179
172
171
166
175
172
174
170



基本統計量の算出

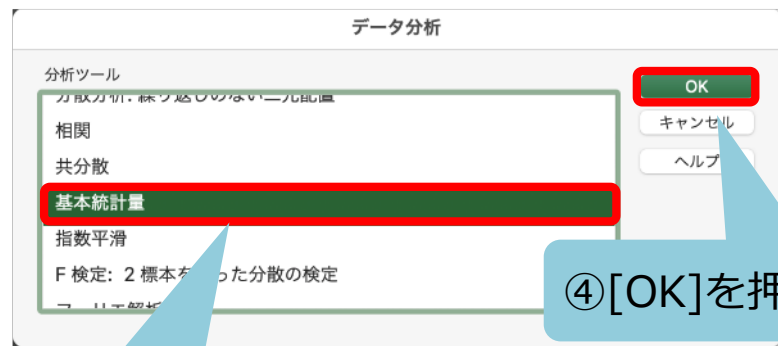
手順① データ分析ツールを起動



①[データ]をクリック

②[データ分析]をクリック

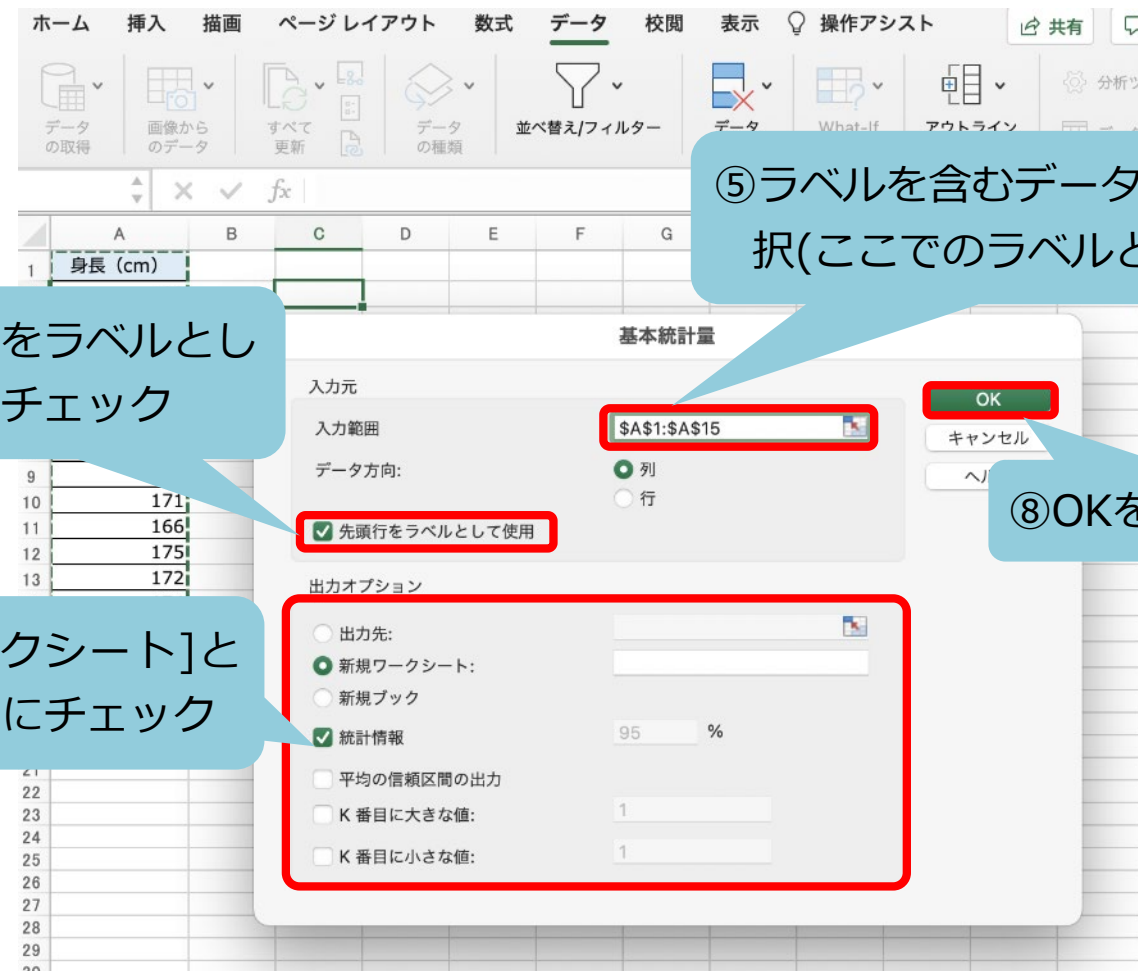
手順② ダイアログが表示されたら、[基本統計量]を選択



③[基本統計量]を選択

④[OK]を押す

手順③ [入力範囲] に基本統計量を算出するデータ範囲であるセルA1～A15を選択し、出力オプションを設定



⑤ラベルを含むデータの範囲を選択(ここでのラベルとはA1)

⑥[先頭行をラベルとして使用]にチェック

⑦[新規ワークシート]と[統計情報]にチェック

⑧OKをクリック

基本統計量の算出

4) 算出結果が出力される

	A	B	C
1	身長(cm)		
2			
3	平均	172.285714	
4	標準誤差	1.34421255	
5	中央値(メジ)	172	
6	最頻値(モー)	175	
7	標準偏差	5.02958282	
8	分散	25.2967033	
9	尖度	0.45209596	
10	歪度	0.41742602	
11	範囲	19	
12	最小	164	
13	最大	183	
14	合計	2412	
15	標本数	14	
16			
17			

【補足】

基本統計量は、関数で算出することも可能
(以下の数式を、値を出したいセルに入力)

平均 = AVERAGE(データ範囲)

中央値 = MEDIAN(データ範囲)

最頻値 = MODE (データ範囲)

標準偏差 = STDEV(データ範囲)

分散 = VAR(データ範囲)

尖度 = KURT(データ範囲)

歪度 = SKEW(データ範囲)

最小 = MIN(データ範囲)

最大 = MAX(データ範囲)

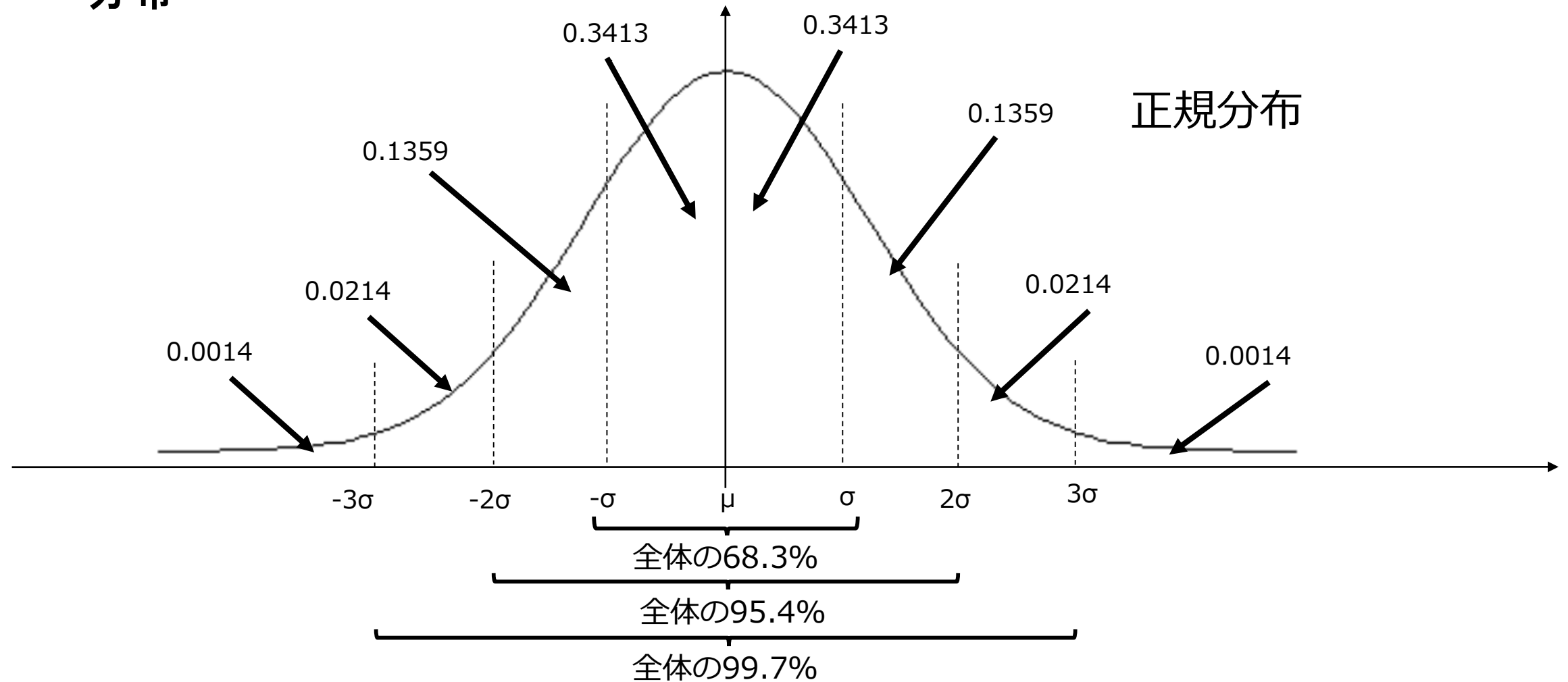
合計 = SUM(データ範囲)

標本数 = COUNT(データ範囲)

- 平均、中央、最頻値 ⇒概ね170cm台前半に集中
- 標準偏差 ⇒仮に正規分布であった場合、167～177cmの間に68.3%のサンプルが入る
- 歪度 ⇒ピークが左に偏る

正規分布とは

ばらつきを伴う事象の分布として、自然界などで最もよく観測される分布



平均を μ 、標準偏差を σ としたとき、
区間 $(\mu - \sigma, \mu + \sigma)$ に入る確率は68.3%、区間 $(\mu - 2\sigma, \mu + 2\sigma)$ に入る確率は95.4%

度数分布表とヒストグラムの描画

- 同じ身長データから度数分布表を作成し、ヒストグラムを描く

手順① Excelで元データに加えて階級間隔を指定する表を用意
基本統計量において最小値164cm、最大値183cmのため、
今回は階級間隔を160cm～185cmの間で5cm刻みとして設定

元データ

	A	B	C
1	身長 (cm)		身長階級 (cm)
2	183		160
3	167		165
4	175		170
5	164		175
6	174		180
7	170		185
8	179		
9	172		
10	171		
11	166		
12	175		
13	172		
14	174		
15	170		
16			
17			

①階級間隔の表を追加。
このとき、数値は上限値を示すことに注意。

例) 165
⇒160超～165以下 を示す

度数分布表とヒストグラムの描画

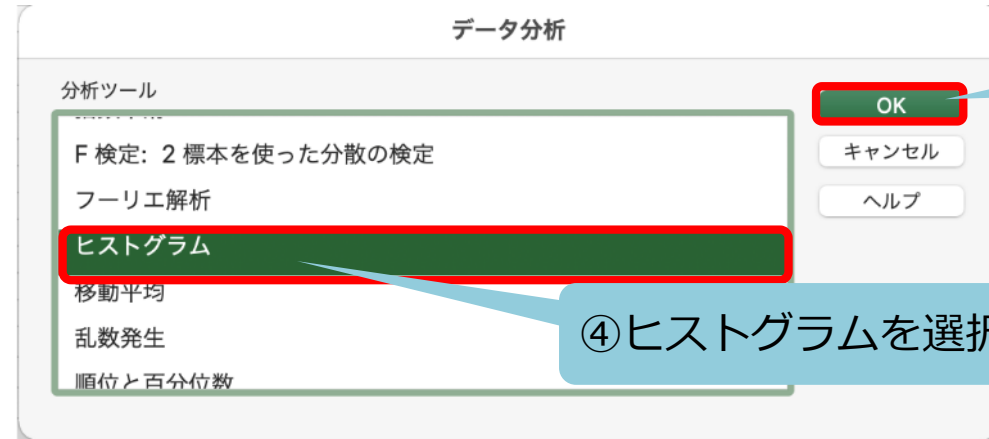
手順② データ分析ツールを起動



②[データ]をクリック

③[データ分析]をクリック

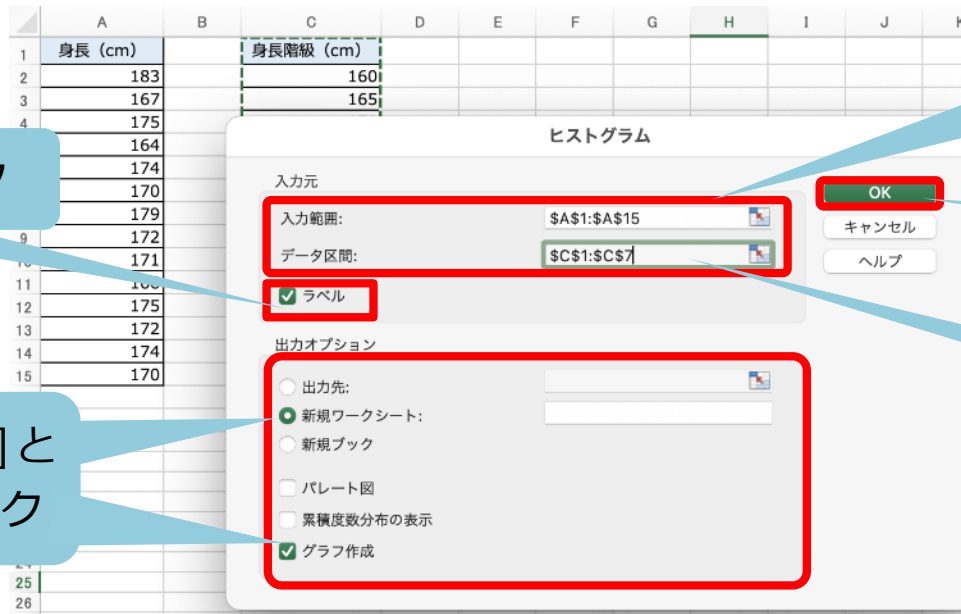
手順③ ヒストグラムを選択



⑤OKをクリック

④ヒストグラムを選択

手順④ [入力範囲] にデータ範囲であるセルA1～A15を、[データ区間] に階級間隔であるC1～C7を選択し、出力オプションを設定



⑧ [ラベル] にチェック

⑨ [新規ワークシート]と
[グラフ作成]にチェック

⑥ラベルを含むデータの範囲
を選択

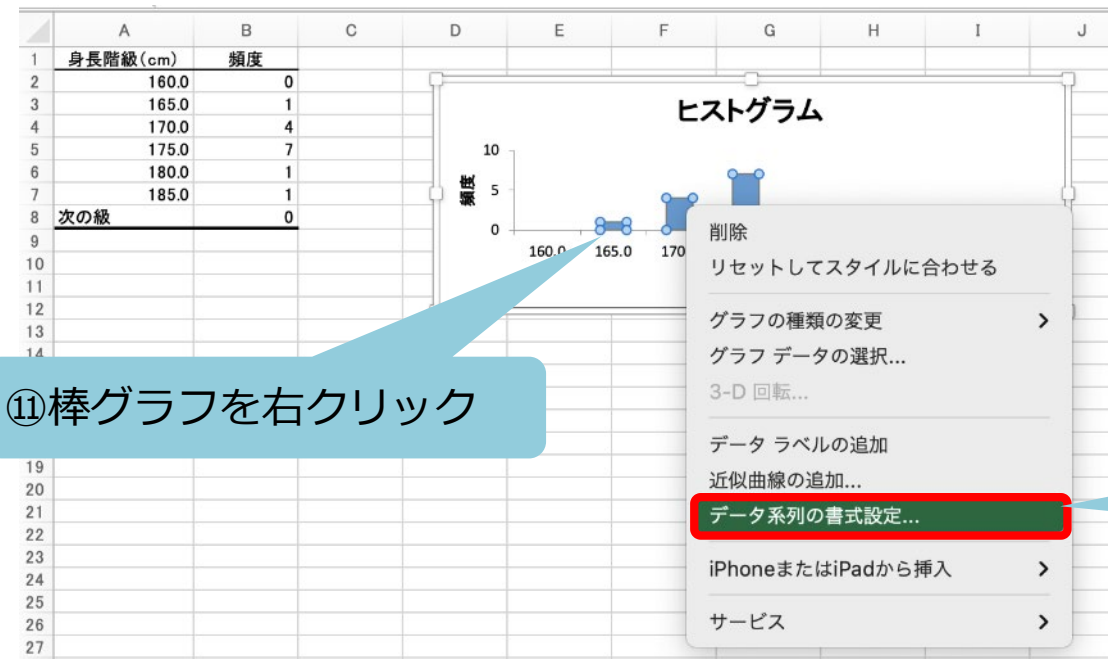
⑩OKをクリック

⑦階級間隔の範囲を選択

度数分布表とヒストグラムの描画

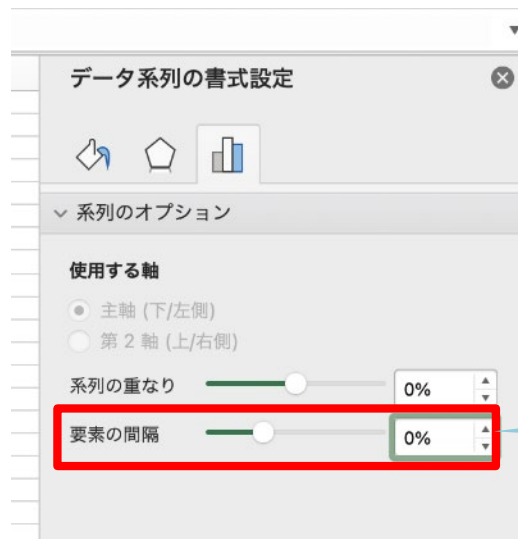
手順⑤ 度数分布表とヒストグラムが同時に出力

手順⑥ ヒストグラムのグラフを右クリックして[データ系列の書式設定]を選択



⑪ 棒グラフを右クリック

⑫ [データ系列の書式設定]をクリック

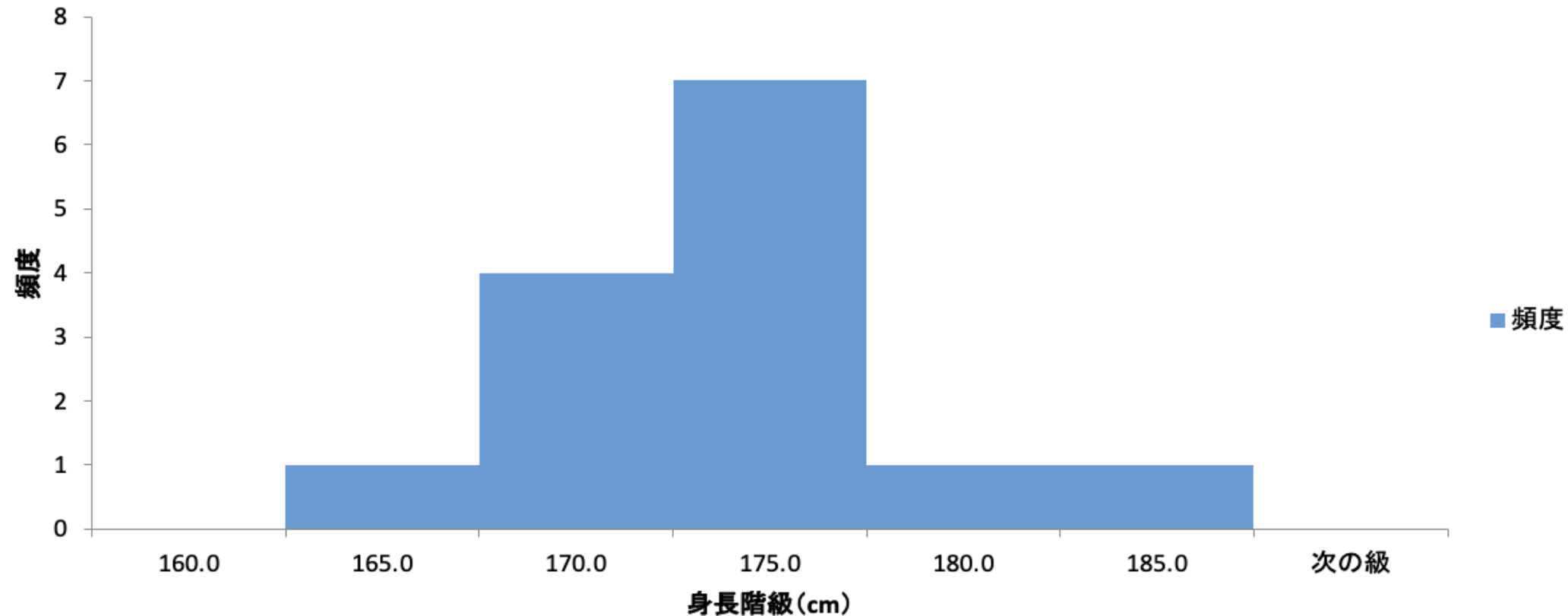


手順⑦ データ系列の書式設定で[要素の間隔]を0%設定する。

⑬ [要素の間隔]を0%にする

度数分布表とヒストグラムの描画

手順⑧ ヒストグラムが完成



- データが集中しているの（ピーク）は170cm超～175cm以下
- データは160cm超～185cm以下の範囲内
- データは175cm以下に偏る

度数分布表とヒストグラムの描画

- 1変数の状況把握のために、ヒストグラムを用いる
- ヒストグラムの元になる表を度数分布表という

度数分布表

身長階級(cm)	頻度
160.0	0
165.0	1
170.0	4
175.0	7
180.0	1
185.0	1
次の級	0

ヒストグラム

