

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：橋本 武彦

第2週の内容紹介

1	● 第2週のまとめ
2	● 「分析ツール」の有効化
3	● 基本統計量の算出
4	● 度数分布表とヒストグラムの描画

※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® 2013 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。

1.第2週のまとめ

Analysisの具体的手法

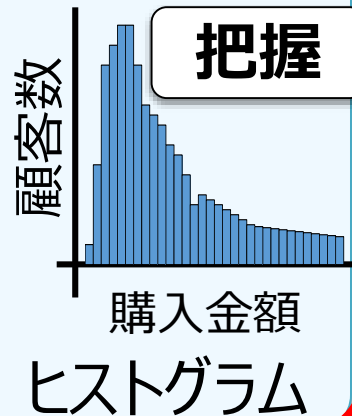
分布 / 比較 / 傾向

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



比較

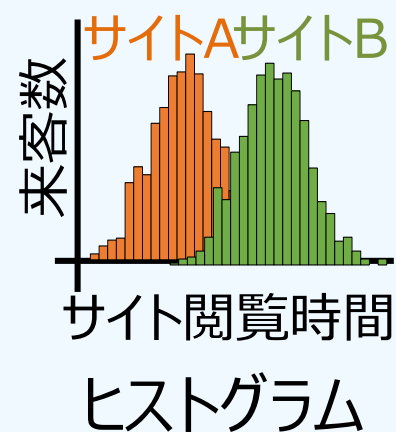
名義 VS 名義

来客数	男	女
サイトA	18	3
サイトB	4	16

クロス集計

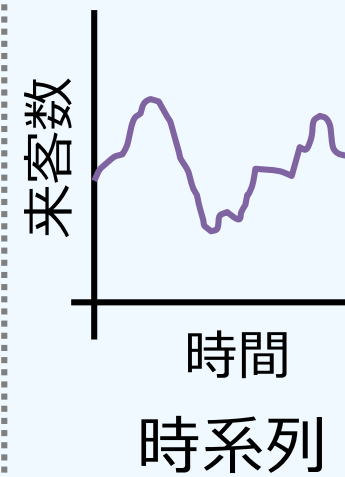
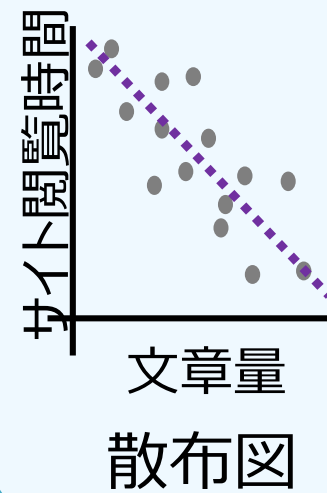
名義 VS 連続

調査



傾向

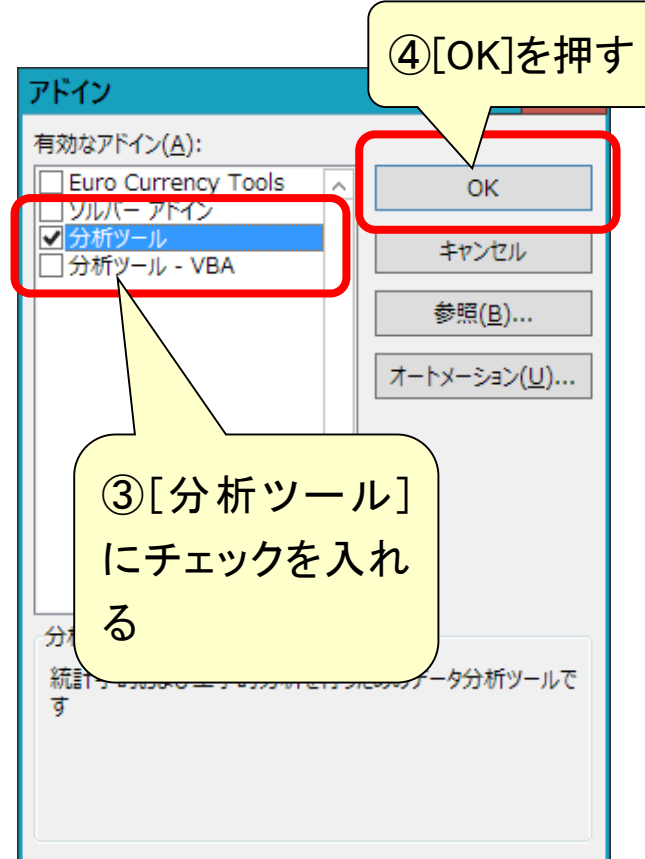
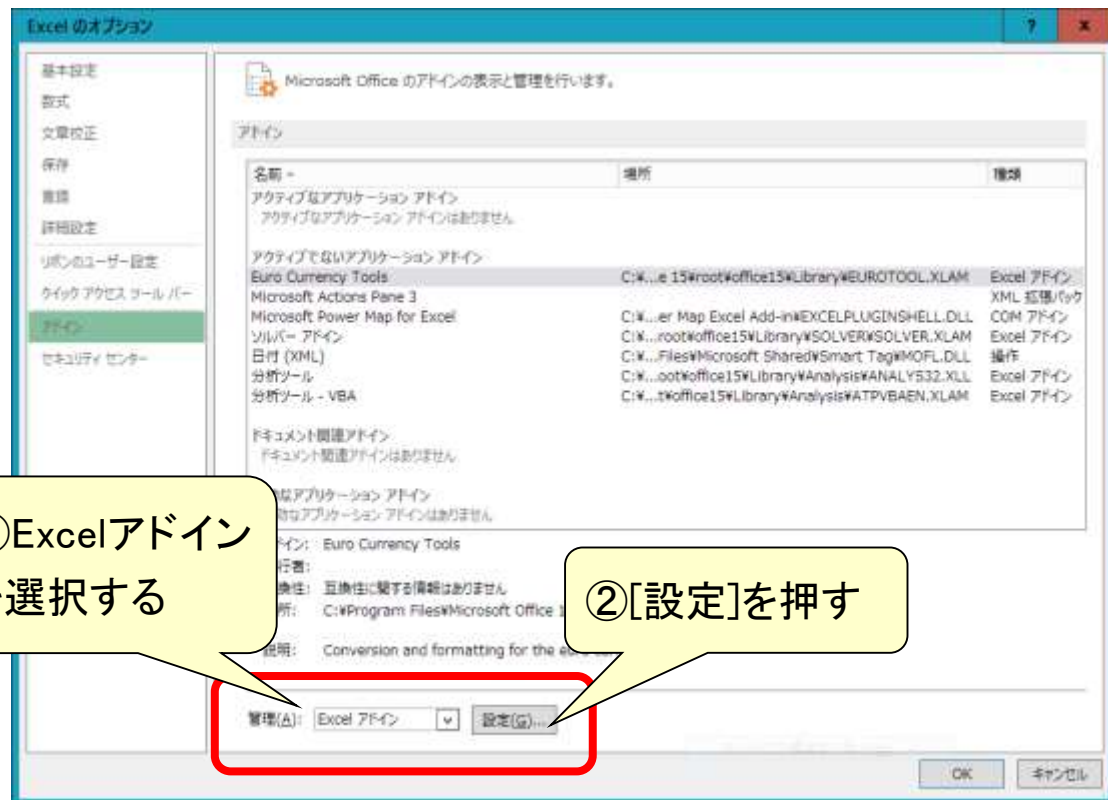
連続 VS 連続



2. 分析ツールの有効化

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]ボタンを押下

2) [分析ツール]にチェックを入れて、[OK]を押下。
これで、分析ツールアドインを有効化



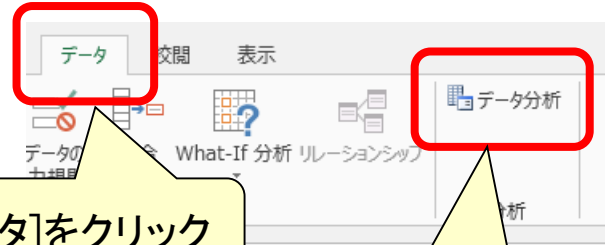
3.基本統計量の算出

- 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。

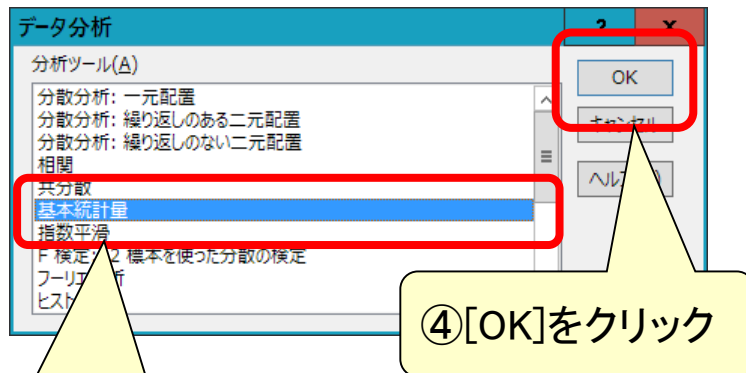
身長 (cm)
183
167
175
164
174
170
179
172
171
166
175
172
174
170

3.基本統計量の算出

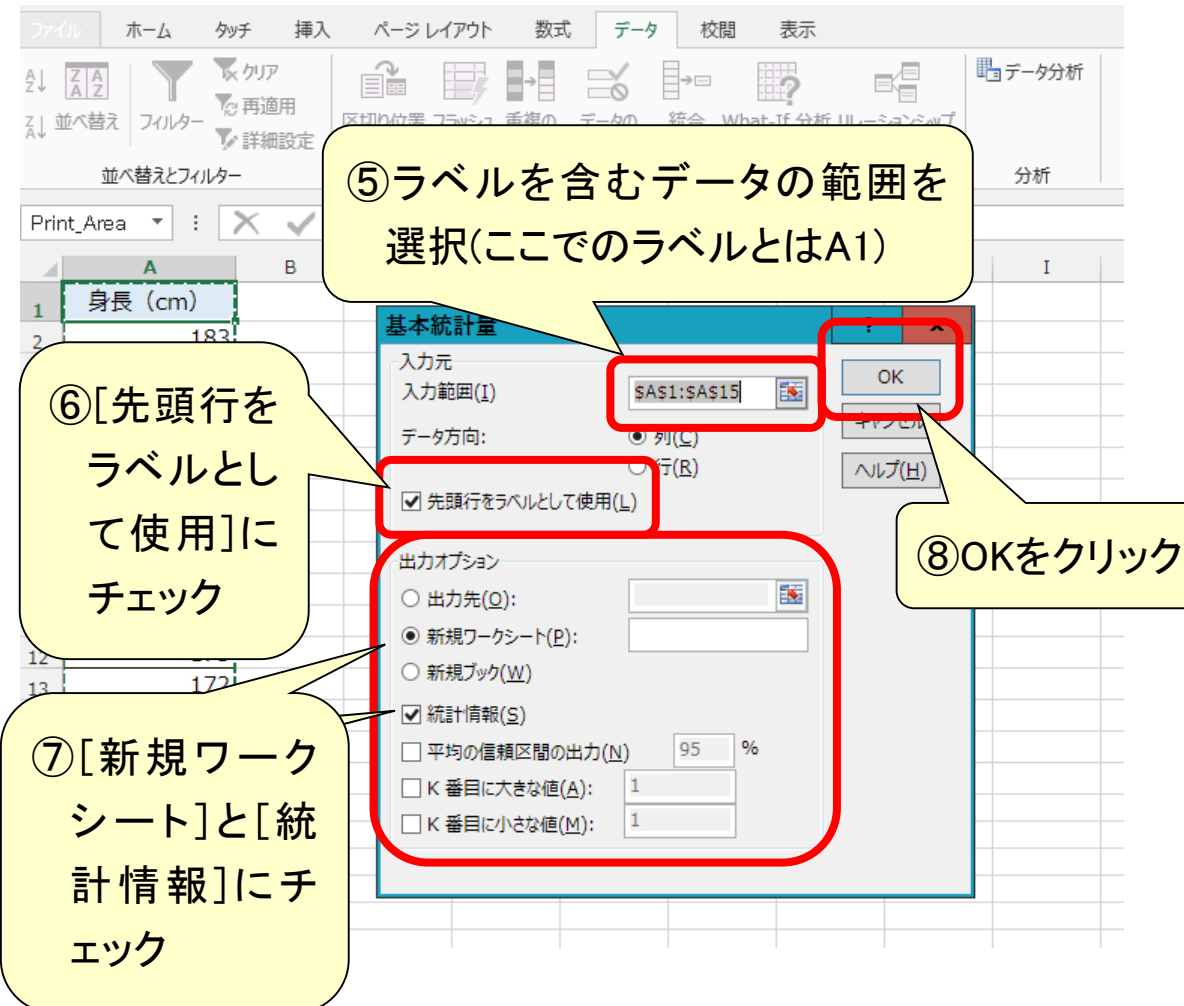
1) データ分析ツールを起動



2) ダイアログが表示されたら、[基本統計量]を選択






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



3.基本統計量の算出

4) 算出結果が出力

A3	:				平均
	A	B	C		
1	身長(cm)				
2					
3	平均	172.2857143			
4	標準誤差	1.34421255			
5	中央値(メジアン)	172			
6	最頻値(モード)	175			
7	標準偏差	5.029582815			
8	分散	25.2967033			
9	尖度	0.452095961			
10	歪度	0.417426015			
11	範囲	19			
12	最小	164			
13	最大	183			
14	合計	2412			
15	標本数	14			
16					

【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 ⇒167~177cmの間に71%のサンプルが入る
- 歪度 ⇒ピークが左に偏る

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、ヒストグラムを描いてみましょう。

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

	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		

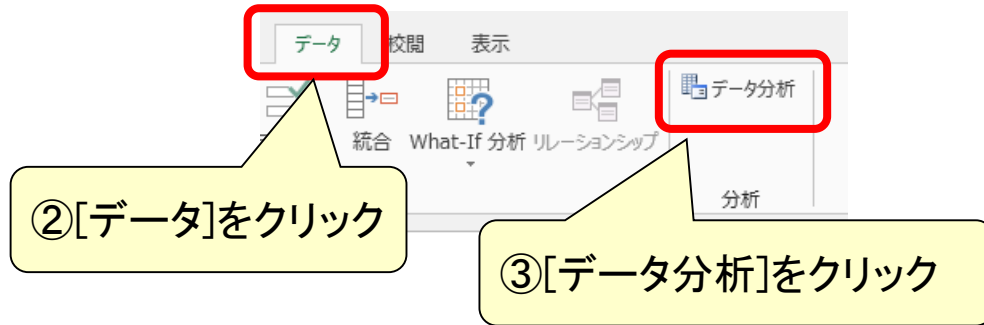
元データ

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。

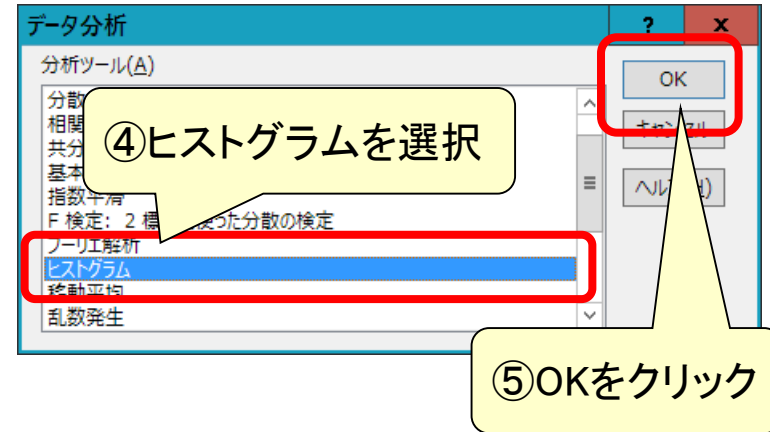
例) 165
⇒160超～165以下 を示す

4. 度数分布表とヒストグラムの描画

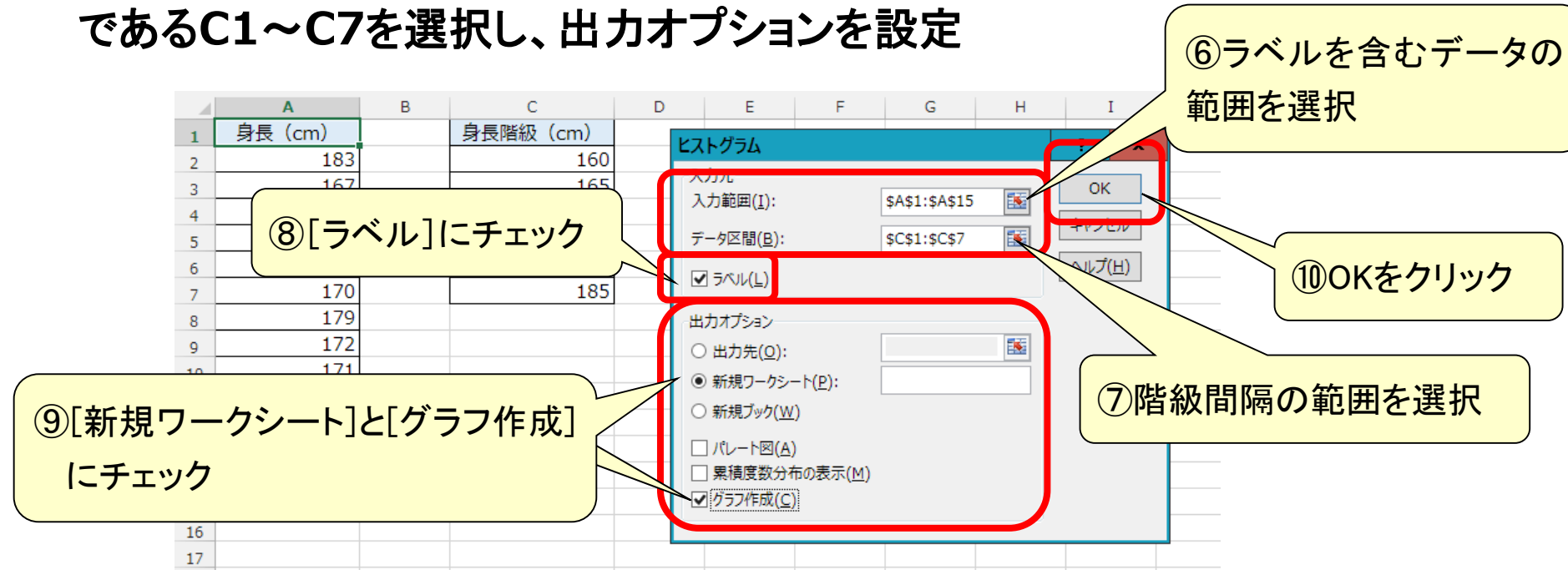
2) データ分析ツールを起動



3) ヒストグラムを選択



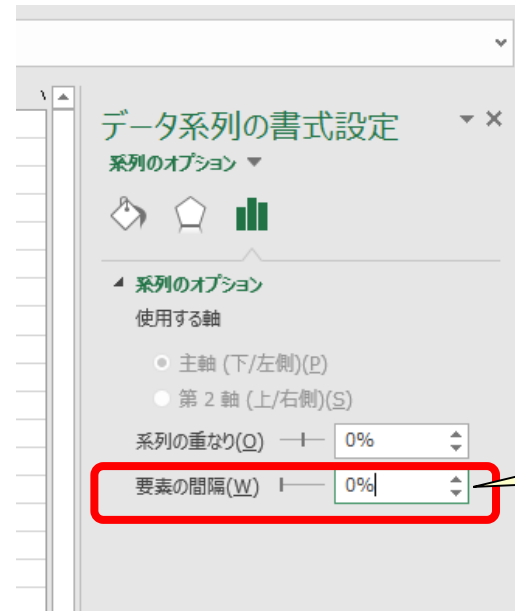
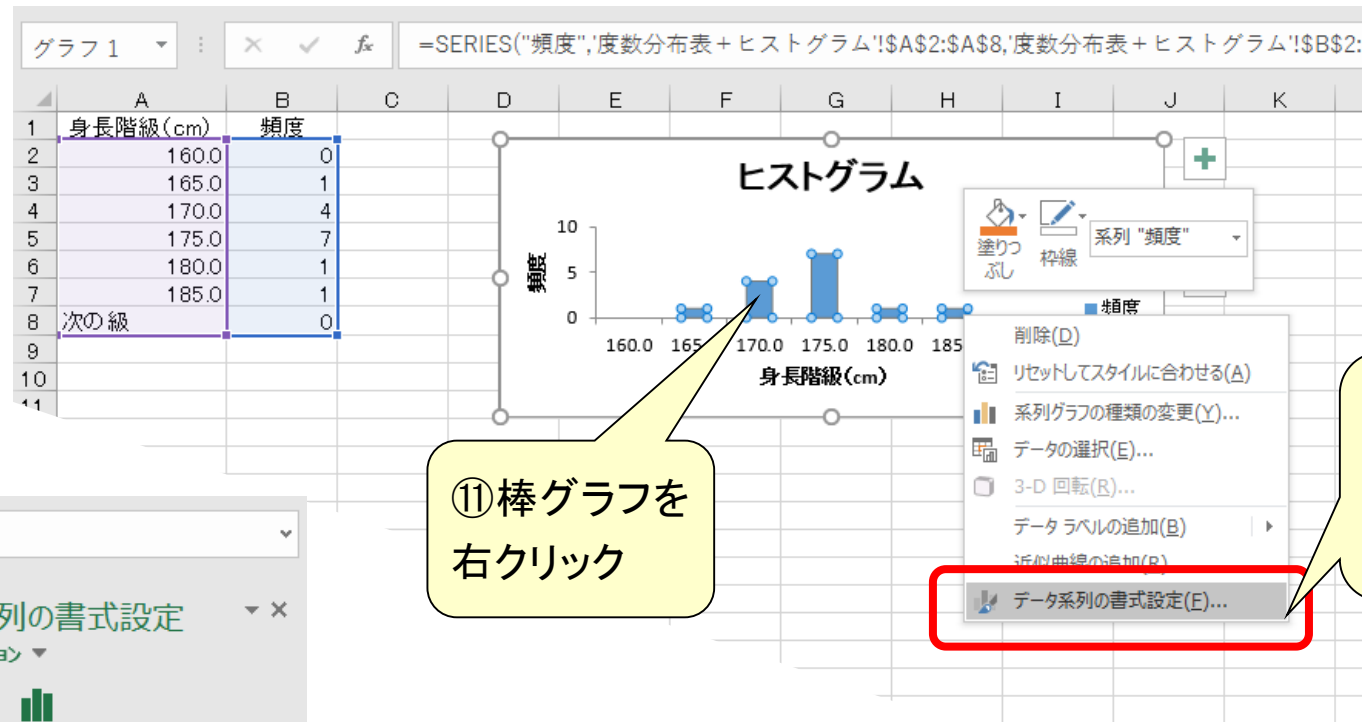
4) [入力範囲]にデータ範囲であるセルA1~A15を、[データ区間]に階級間隔であるC1~C7を選択し、出力オプションを設定



4. 度数分布表とヒストグラムの描画

4) 度数分布表とヒストグラムが同時に出力

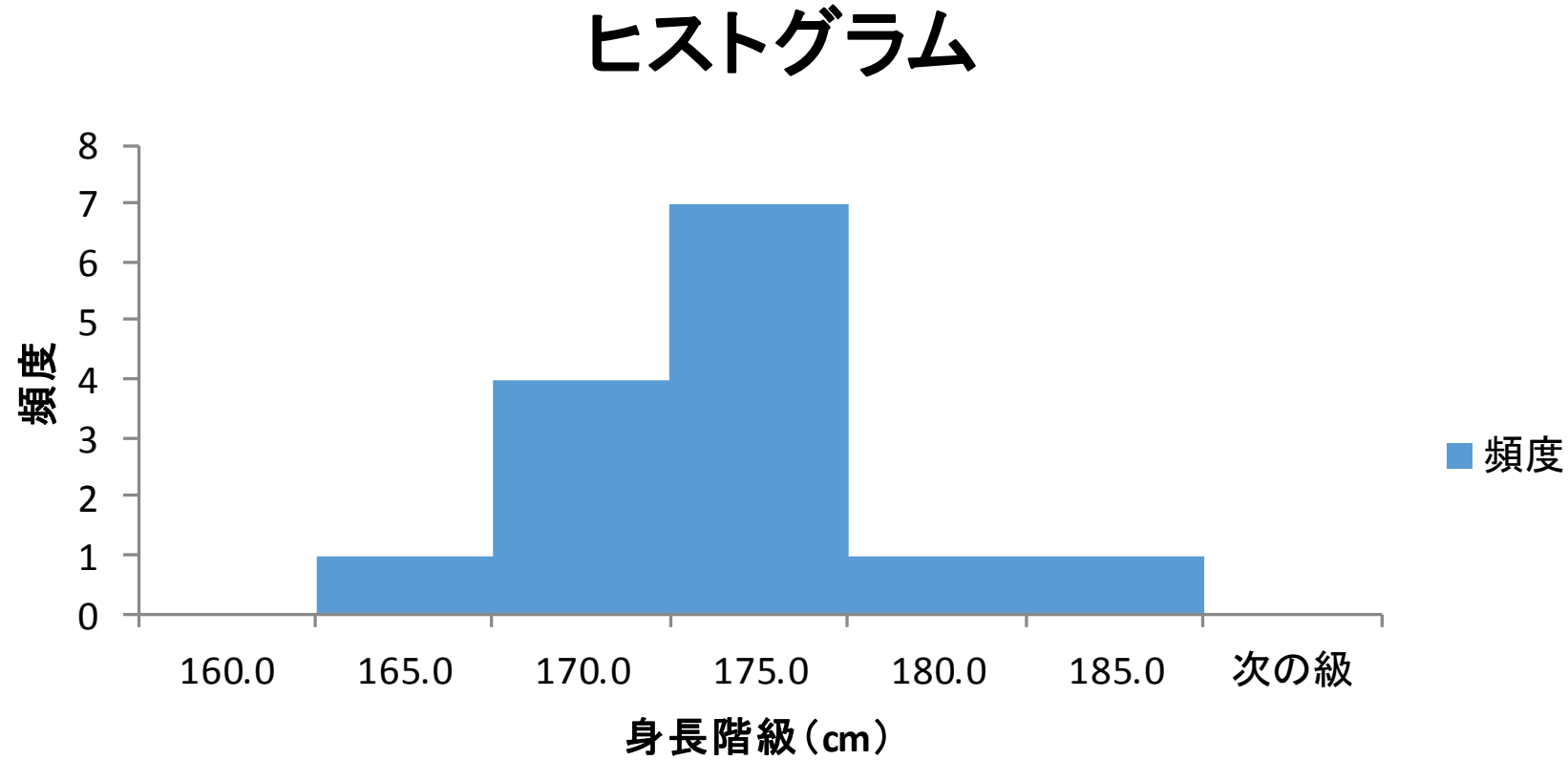
5) ヒストグラムのグラフを右クリックして[データ系列の書式設定]を選択



6) データ系列の書式設定で[要素の間隔]を0%設定する。

4.度数分布表とヒストグラムの描画

7) 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

4.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

