統計学の基本

統計学の基礎① ビックデータとはなにか・Excel基本操作

観光ビッグデータとは？

観光ビッグデータ

- 観光ビッグデータの明確な定義はない。
- 観光客の観光行動を反映する大規模・多種・複数情報源由来のデータ群を言うことが多い。

具体例

- スマートフォン等を使用してネットワークに接続したログ
- コンテンツのアクセスログやダウンロード数
- ソーシャルネットワーキングサービス（SNS）への投稿
- GPS機能による位置情報
- 自動車のプローブ情報
- 交通事業者の乗降の情報
- 観光施設の日別の利用者数

観光ビッグデータ分析の必要性

Before

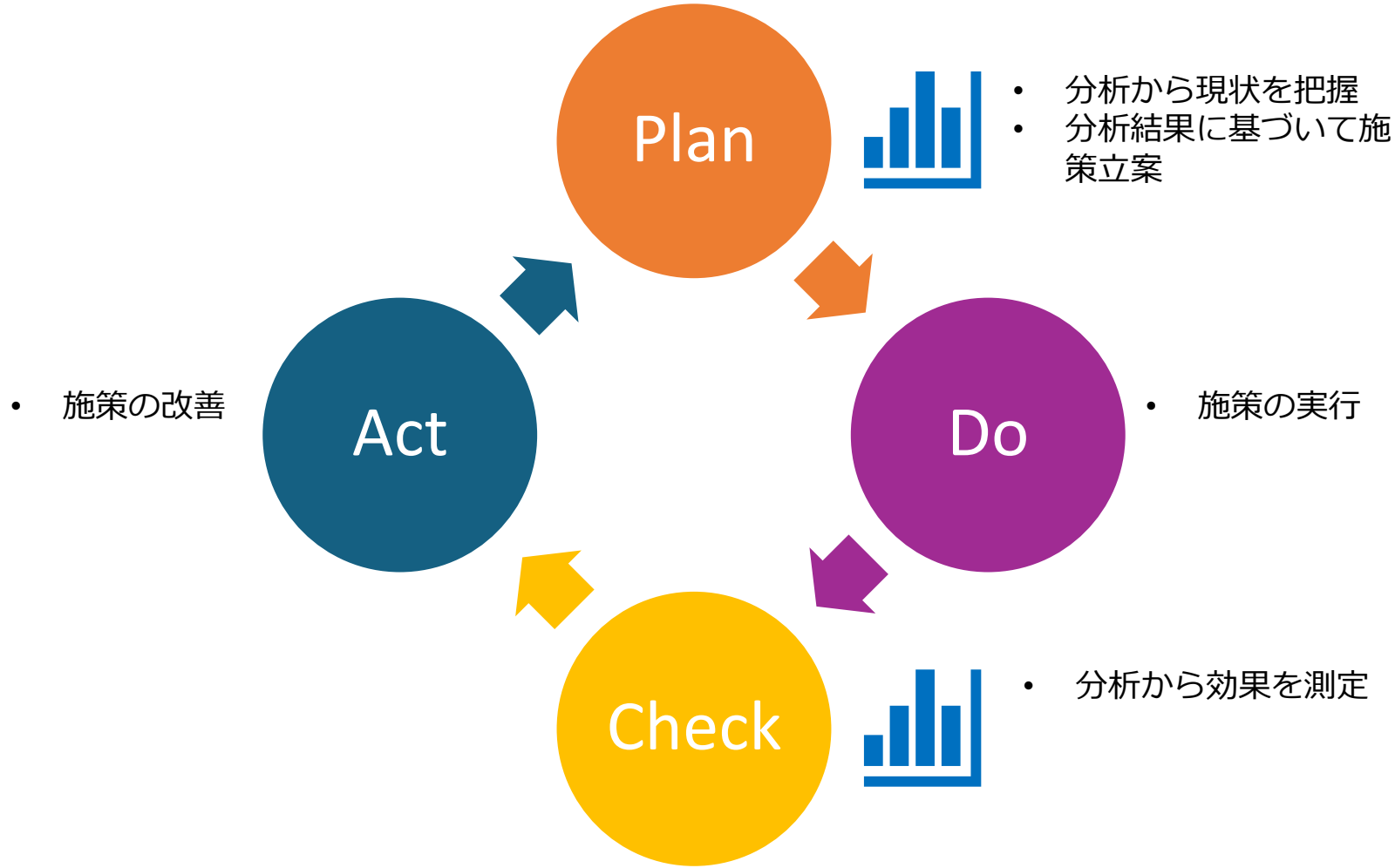
- 経験者の暗黙知
- 目の前のできごとをベースとした施策

After

- 事実に基づく明確なコンセプトに基づいた戦略（ブランディング）の策定
- KPIの設定・PDCAサイクルの確立
- プロモーション

分析データの活用

PDCAサイクルに位置づけて活用していくことが重要



分析における課題

観光ビッグデータは、それぞれのデータの精度は公的統計ほどに高くないため、分析の蓄積によって精度が高まっていく。

ミスリード防止

- 観光ビッグデータは、データにより特性やサンプルサイズなどの情報が異なるため、単純な比較や推測はできない。
- 分析結果等をミスリードしないよう、各データの特性や情報源の明示、前提条件や注釈の記載などより一層の慎重な配慮が必要

組み合わせ

- 観光ビッグデータではカバーされない情報も多い。
- 団体などにおいて分析を行う場合には、分析の目的に応じて、既存の公的統計やその他の調査などを組み合わせた分析が必要。

フィードバック

- データ分析によって立案された観光施策を効果的に展開するためには、施策実施後の結果を分析にフィードバックする。
- フィードバックに基づいて分析の精度を高めることが重要。

データの整理と分析

- 観光プロモーションに向けた検討を行うためには、ターゲットの細分化が必要。
- セグメントについても細分化とそれを可能にするデータの整理と分析が期待される。

観光分野におけるビッグデータの効果的な活用

動機の明確化	データの見方	利用するデータ	具体的な視点
• 地域の課題の発見 • 解決方法の検討 • 各種施策などの効果測定	• 思い込みの排除 • 地域の切実な思いと肌感覚 • データには表れない情報の補完 • 現場の視点に基づく取り組み	• データの不完全さを認識する • 確かな情報との整合性や対応関係 • 既存の統計調査等との組み合わせ	• 国・地域、時間帯、季節、日本人観光客との比較 • 類似した状況を参考にする • 周辺地域の特徴や強みの活用 • 俯瞰する視点

まとめ

- 観光ビッグデータ分析を行うことで、事実に基づく観光施策を打ち出すことができる。
- 分析はPDCAサイクルに位置づけてブラッシュアップしていくことが重要。
- 観光ビッグデータの弱点を理解し、各種の視点から注意深く取り扱う必要がある。
- 観光ビッグデータを効果的に活用するために必要な視点を身につけよう。

ワーク

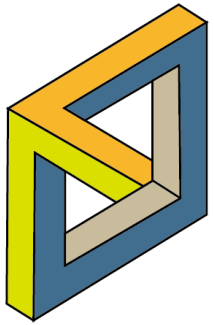
データ分析レベル診断

あなたのデータ分析レベルの診断をしましょう。

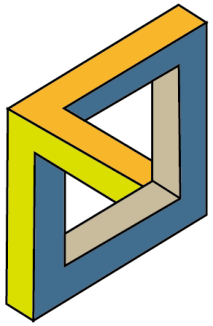
質問	回答
身の回りに分析できるデータがある。	
統計を学んだことがある。	
Excelを使ったことがある。	
与えられた課題に対して分析をしたことがある。	
分析を設計したことがある。	
分析結果を元に施策を考えたことがある。	

データ分析のアイディア

- あなたがデータ分析を用いて解決したい課題を考えてみましょう。
- また、その課題を解決するために必要なデータにはどのようなものがあるかを考えましょう。



ハンズオン



データ分析アドインとは？

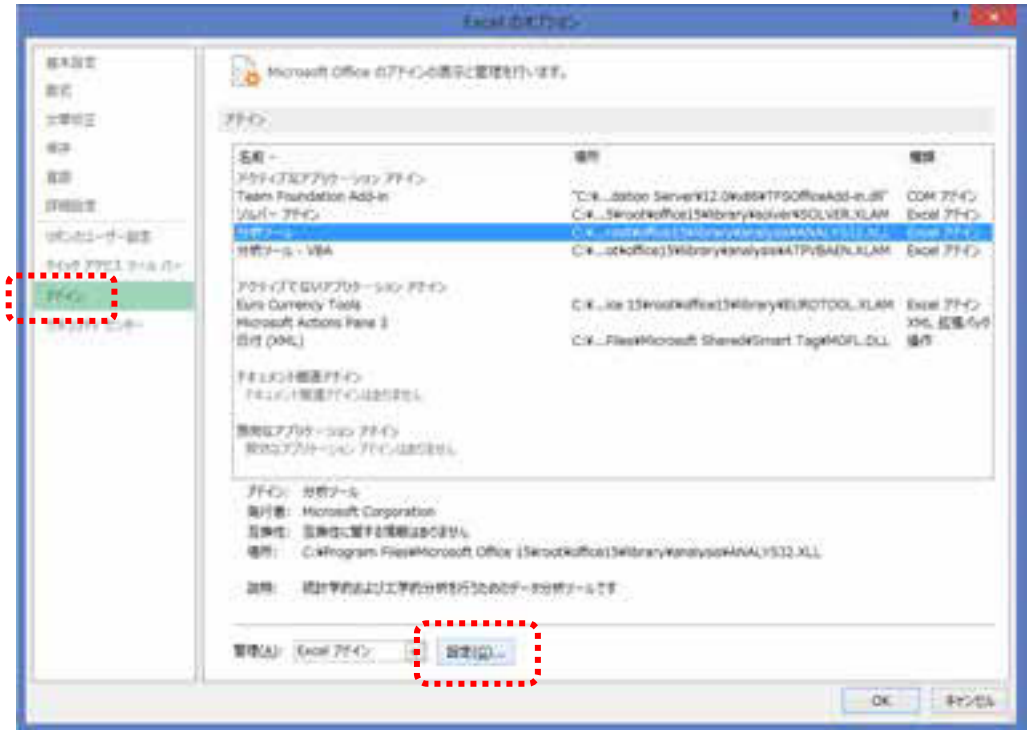
- データ分析や統計学でよく行われる分析パターンをまとめたツール。
  複雑な分析に対して、データとパラメータを指定すると素早く簡単に結果を得ることができる。
- 利用するためには「アドインのインストール」が必要。

分析ツールのインストール（１）



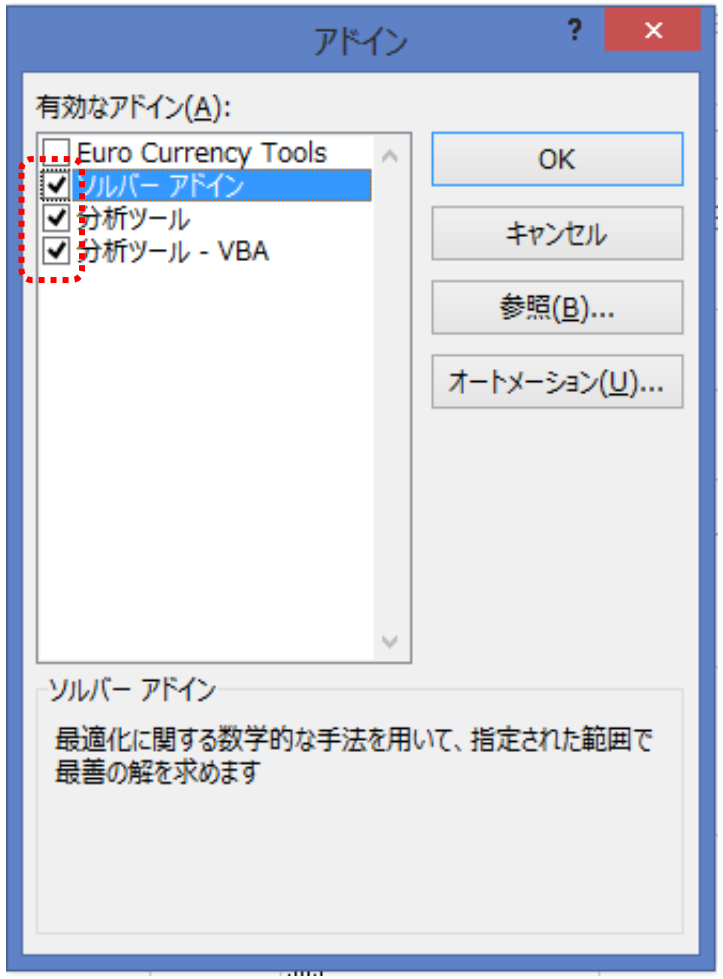
[ファイル]を選択して、メニュー画面が表示されたら、[オプション]を選択する。

分析ツールのインストール（２）



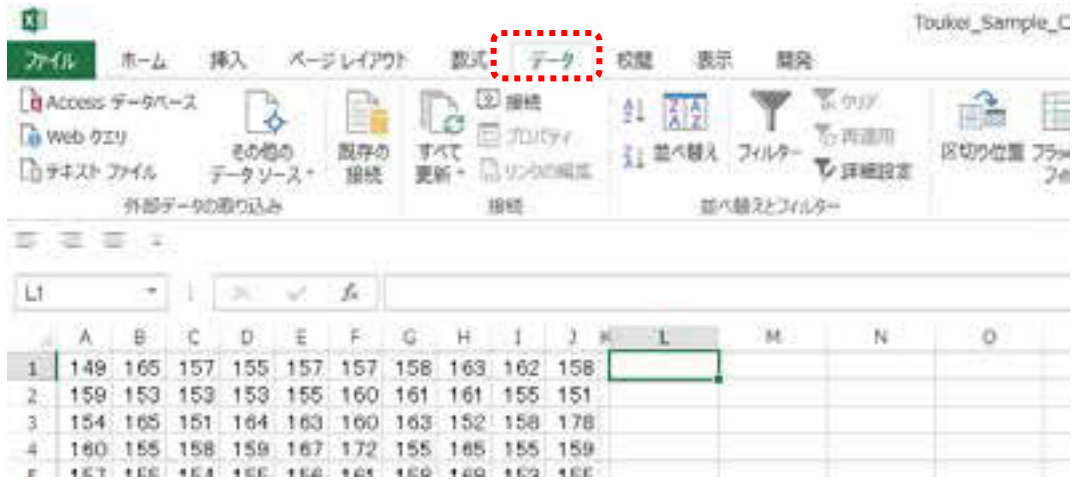
[アドイン]を選択して、
[設定]をクリックする。

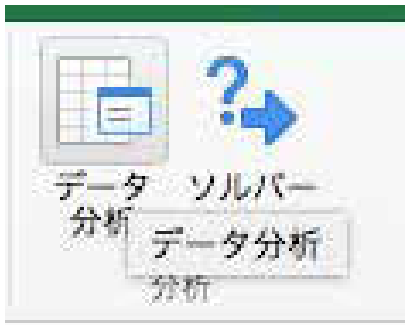
分析ツールのインストール（３）



[アドイン]の項目が出たら、[分析ツール]のチェックボックスにチェックを入れて[OK]を選択する。

分析ツールのインストール（４）





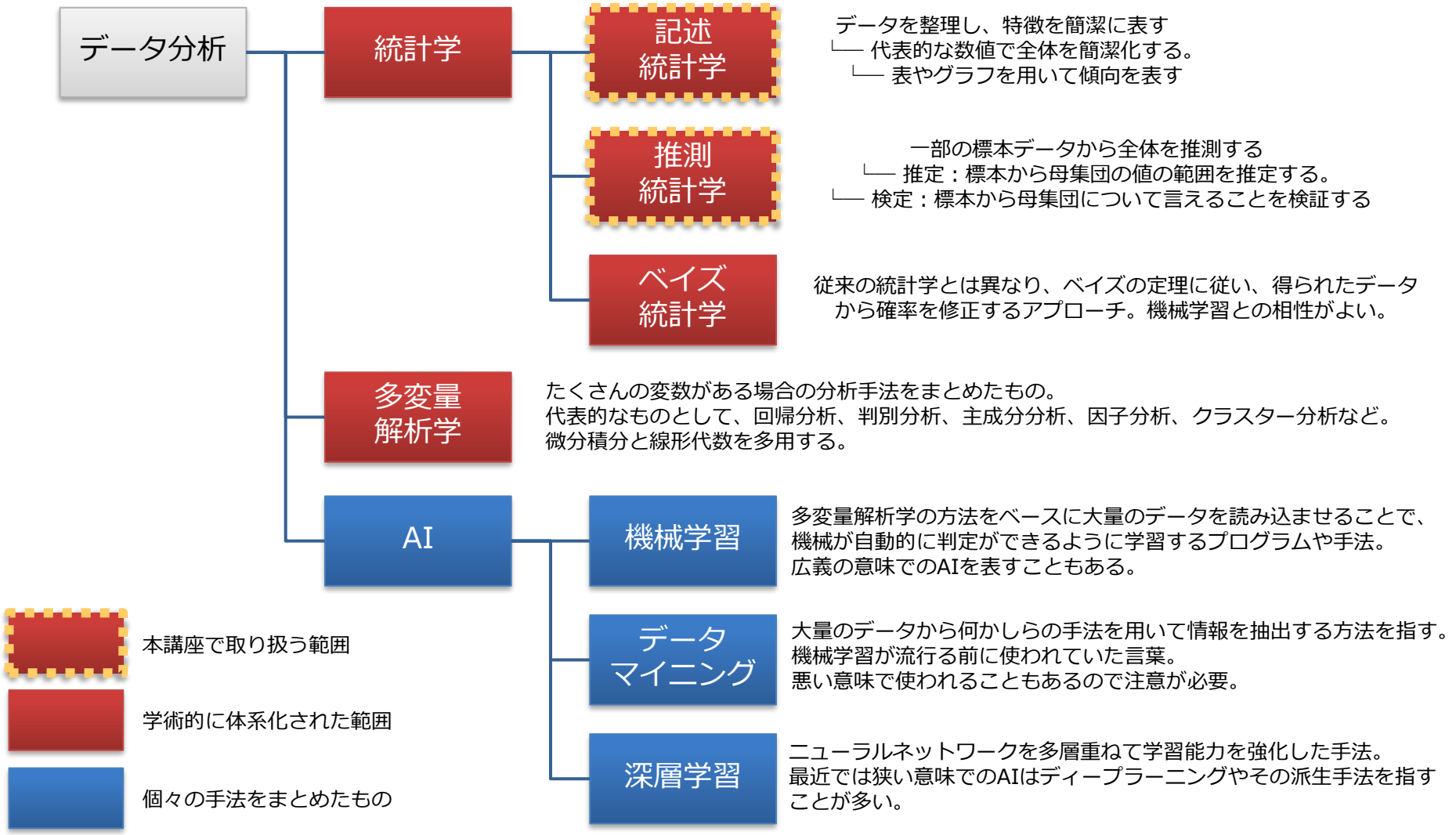
Excel画面の上部[データ]のタブを選択し、画面の右端に[データ分析]という項目が出たらインストール完了です。

統計学の基本

統計学の基礎　②データの種類

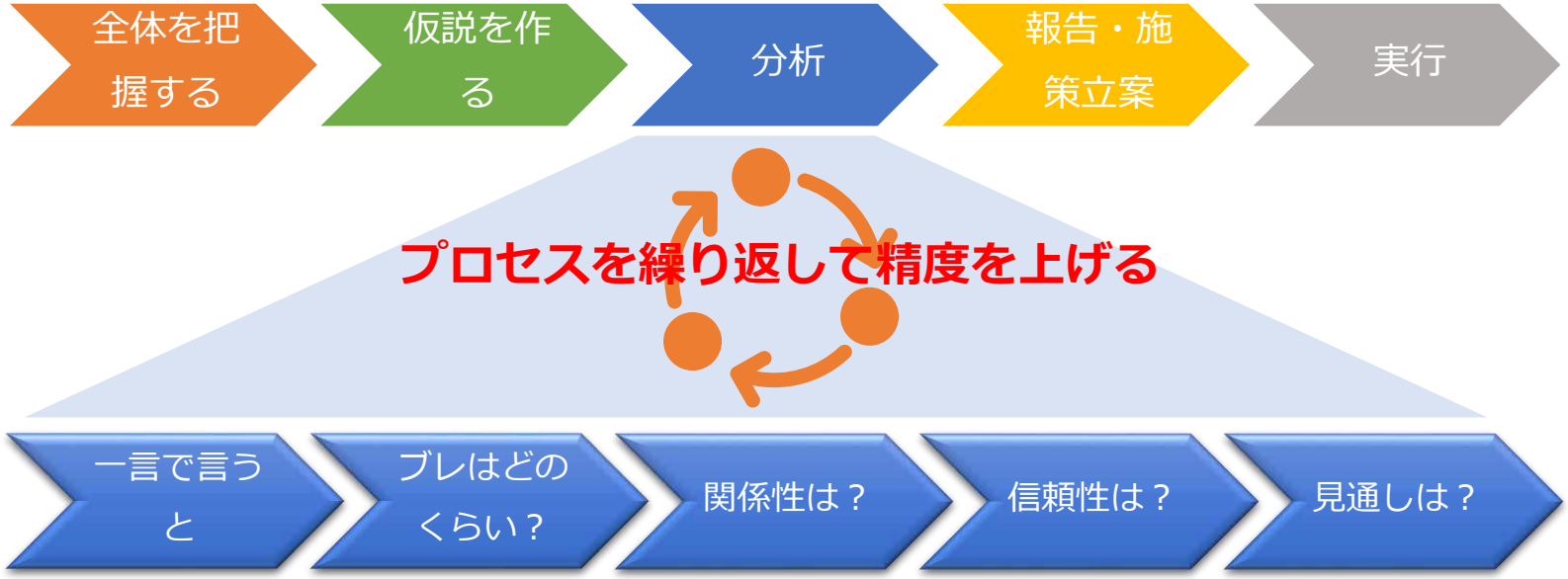
統計学・データ分析マップ

- 「データ分析」の手法は様々なものがある。知識を整理して理解する。



データ分析の流れ

- データ分析もビジネスも「言うべきこと・やるべきこと」は同じ！



- 代表値
  - 平均値
  - 中央値
  - 最頻値
- その他
  - 歪度
  - 尖度

- 散布度
  - 平均偏差
  - 分散
  - 標準偏差
- 範囲
  - 最大値
  - 最小値
  - 四分位数

- 相関係数
- クロス表
- 相関比

- 仮説検定

- 回帰分析

データの特徴

- データは主に、そのデータが取られた観測対象（ケース）と観測項目（変数）によって表の形式でまとめられていることが多い。これをデータ行列という。
- 統計分析が必要となるデータは、観測対象ごとにばらつきがあることが特徴となる。

観測項目
（変数）

観測対象
（ケース）

	A	B	C	D	E	F	G	H	I	J
1	#	住宅価格	坪単価	坪数	建蔽率	建物単価	施主年齢	施主生年	販売会社	都道府県
2	1	47,400,000	563,000	44.69	0.834	482,000				
3	2	60,300,000	694,000	41.74	0.759	518,000				
4	3	53,700,000	744,000	42.89	0.955	384,000				
5	4	29,300,000	613,000	37.87	0.892	440,000				
6	5	56,200,000	717,000	38.87	1.005	504,000			1977	ダイワハウス 東京都
7	6	45,100,000	662,000	40.31	0.734	534,000	40	1976	中央住宅	千葉県
8	7	44,700,000	648,000	42.87	0.849	516,000	37	1979	中央住宅	埼玉県
9	8	66,200,000	750,000	36.49	0.834	396,000	43	1973	中央住宅	東京都
10	9	47,500,000	793,000	30.22	0.86	469,000	38	1978	ダイワハウス	埼玉県
11	10	53,200,000	761,000	35.17	0.913	478,000	37	1979	中央住宅	東京都
12	11	38,100,000	583,000	41.98	0.866	415,000	42	1974	中央住宅	神奈川県
13	12	23,700,000	496,000	35.78	0.873	388,000	41	1975	中央住宅	神奈川県
14	13	25,700,000	568,000	40.57	0.904	430,000	39	1977	住友不動産	千葉県
15	14	46,600,000	660,000	39.98	0.843	537,000	39	1977	ダイワハウス	埼玉県
16	15	60,600,000	587,000	34.8	0.844	425,000	45	1971	ダイワハウス	東京都
17	16	63,100,000	672,000	45.08	0.809	466,000	36	1980	住友不動産	東京都
18	17	46,700,000	667,000	43.48	0.882	396,000	41	1975	住友不動産	神奈川県

観測対象ごとにばらついている

【サンプル】不動産価格

データの種類

- 統計解析の対象となっているデータは特徴に応じて以下のような分類をされる。
- はじめに、「質的データ」と「量的データ」の違いを理解しよう。

			性質	例	使用できる演算
データ	質的データ	名義尺度	区別のみ 順序はない	性別、居住地、職業区分、Yes/No	分類
		順序尺度	区別に加えて、 順序がある。 差は取れない。	5段階評価	分類、クラスの併合、累積
	量的データ	間隔尺度	差の大きさに意味がある。比には意味がない。	摂氏温度、知能指数、偏差値、…	足し算や引き算（合計、平均）
		比率尺度	差も比も意味がある。 （絶対ゼロがある）	身長、体重、得点、…	四則演算

ワーク

身の回りのデータの分類

身の回りにあるデータを「データの種類」に分類します。

身の回りにある「データ」をあげていきます。

「データの種類」のどこに分類されるか

個人で考える → グループで意見交換 → 全体共有

ワークの進め方（1）



後で見て分かるように
テーマを書いておきましょう

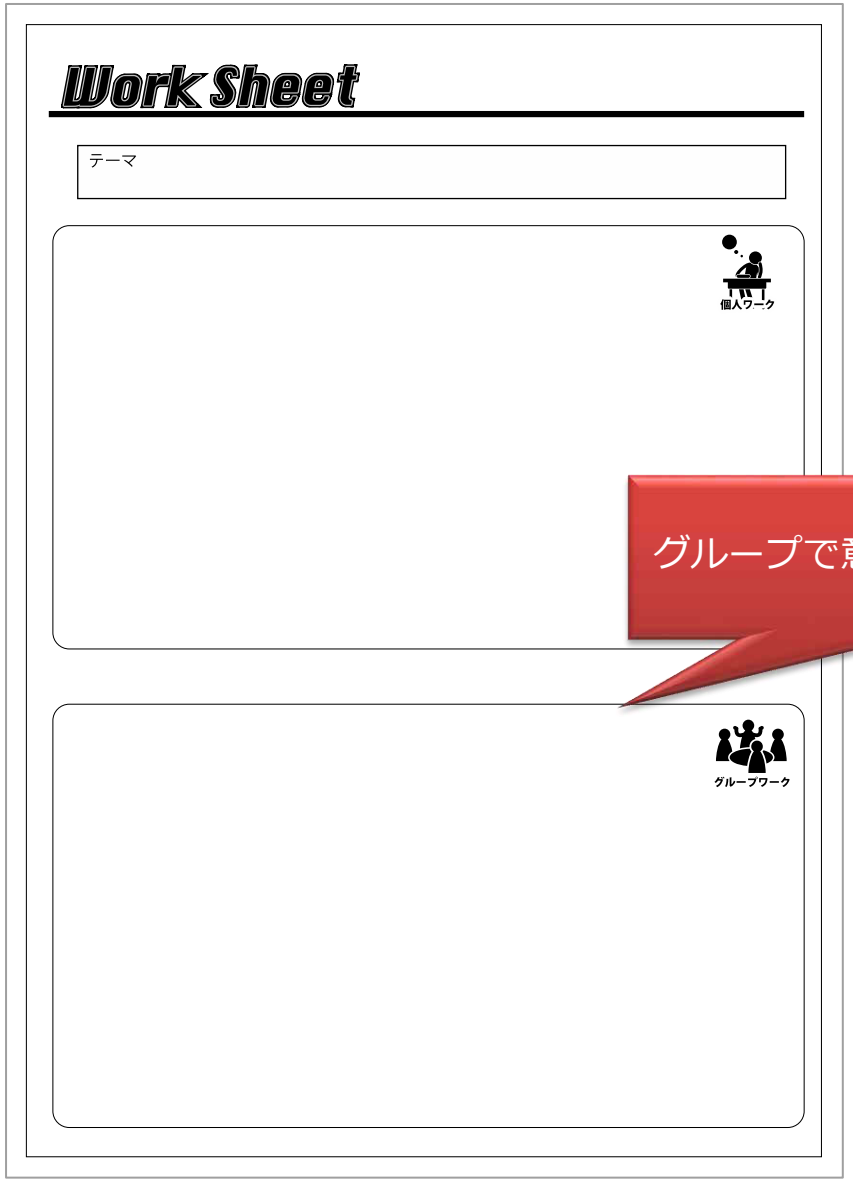
個人の考えを書いておきましょう

身の回りにある「データ」をあげてください。

名義尺度・順序尺度・間隔尺度・比尺度
のどれに当たるか、分類してみましょう。

データ：ここでは数字で表されているものであれば
何でも構いません。

ワークの進め方（２）



グループで意見交換をしましょう。

個人での分類とその理由をグループ内で意見交換しましょう。
グループで議論して、全員が合意する結論を目指しましょう。

ワークの進め方（３）

グループごとの意見を全体に共有しましょう。

名義尺度、順序尺度、間隔尺度、比尺度
それぞれに、どのようなデータを見つけたか教えてください。

※他のグループの説明から新しい視点を見つけてください。



統計学の基本

統計学の基礎　③代表値の算出

ケース

- あなたは、「秋の彩り紅葉日帰りバスツアー」の企画担当となりました。Go to などの補助金制度を利用して、以下のような価格帯で販売をしたところ、定員20名のツアーが3日間満席となりました。

分類	価格
大人	8,000円
小人	2,000円
シニア（65歳以上）	5,000円
乳幼児（小学生未満）	0円
大人（座席指定）	10,000円

日程ごとの申し込み人数

	3月20日		3月21日		3月22日	
#	分類	金額	分類	金額	分類	金額
1	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
2	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
3	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
4	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
5	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
6	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
7	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
8	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
9	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
10	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
11	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
12	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
13	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
14	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
15	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
16	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
17	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
18	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
19	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
20	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0

代表値① 平均値

- 平均値とは

	3月20日		3月21日		3月22日	
#	分類	金額	分類	金額	分類	金額
1	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
2	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
3	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
4	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
5	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
6	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
7	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
8	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
9	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
10	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
11	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
12	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
13	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
14	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
15	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
16	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
17	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
18	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
19	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
20	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
	合計	100,000	合計	100,000	合計	100,000
	平均値	5,000	平均値	5,000	平均値	5,000

EXCEL	=average(範囲) =sum(範囲)/count(範囲)

- 平均値（mean）
  - 「平らに均（なら）す」
  - 一言で要約する場合は平均値を使う
  - 記号「μ（ミュー）」で表される。

<メリット>
- 知名度・わかりやすさ

<デメリット>
- データの「歪み」がある場合に適切に情報を伝えらえないことがある。

代表値② 中央値

- 中央値とは

	3月20日		3月21日		3月22日	
#	分類	金額	分類	金額	分類	金額
1	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
2	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
3	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
4	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
5	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
6	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
7	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
8	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
9	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
10	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
11	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
12	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
13	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
14	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
15	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
16	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
17	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
18	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
19	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
20	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
	合計	100,000	合計	100,000	合計	100,000
	平均値	5,000	平均値	5,000	平均値	5,000
	中央値	5,000	中央値	5,000	中央値	5,000

EXCEL	=median (範囲)

- 中央値（median）
  - 集団の中央に位置する値
  - 大きい（小さい）順に並べて中央に来る値

<メリット>

- データの「歪み」がある場合にも結果が安定する。

<デメリット>

- 計算が面倒（並べ替え・場合分け）
- 知名度が低い

代表値③ 最頻値

- 最頻値とは

	3月20日		3月21日		3月22日	
#	分類	金額	分類	金額	分類	金額
1	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
2	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
3	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
4	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
5	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
6	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
7	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
8	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
9	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
10	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
11	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
12	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
13	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
14	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
15	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
16	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
17	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
18	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
19	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
20	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
	合計	100,000	合計	100,000	合計	100,000
	平均値	5,000	平均値	5,000	平均値	5,000
	中央値	5,000	中央値	5,000	中央値	5,000
	最頻値	5,000	最頻値	8,000	最頻値	5,000
				2,000		

EXCEL	=mode.mult (範囲)

- 最頻値（mode）
  - データの中で、最も出現する頻度が高い値（階級値）

<メリット>
- ボリュームゾーンがわかる
- 最も出現可能性が高い値がわかる

<デメリット>
- 全体の傾向が掴めない

代表値の特徴

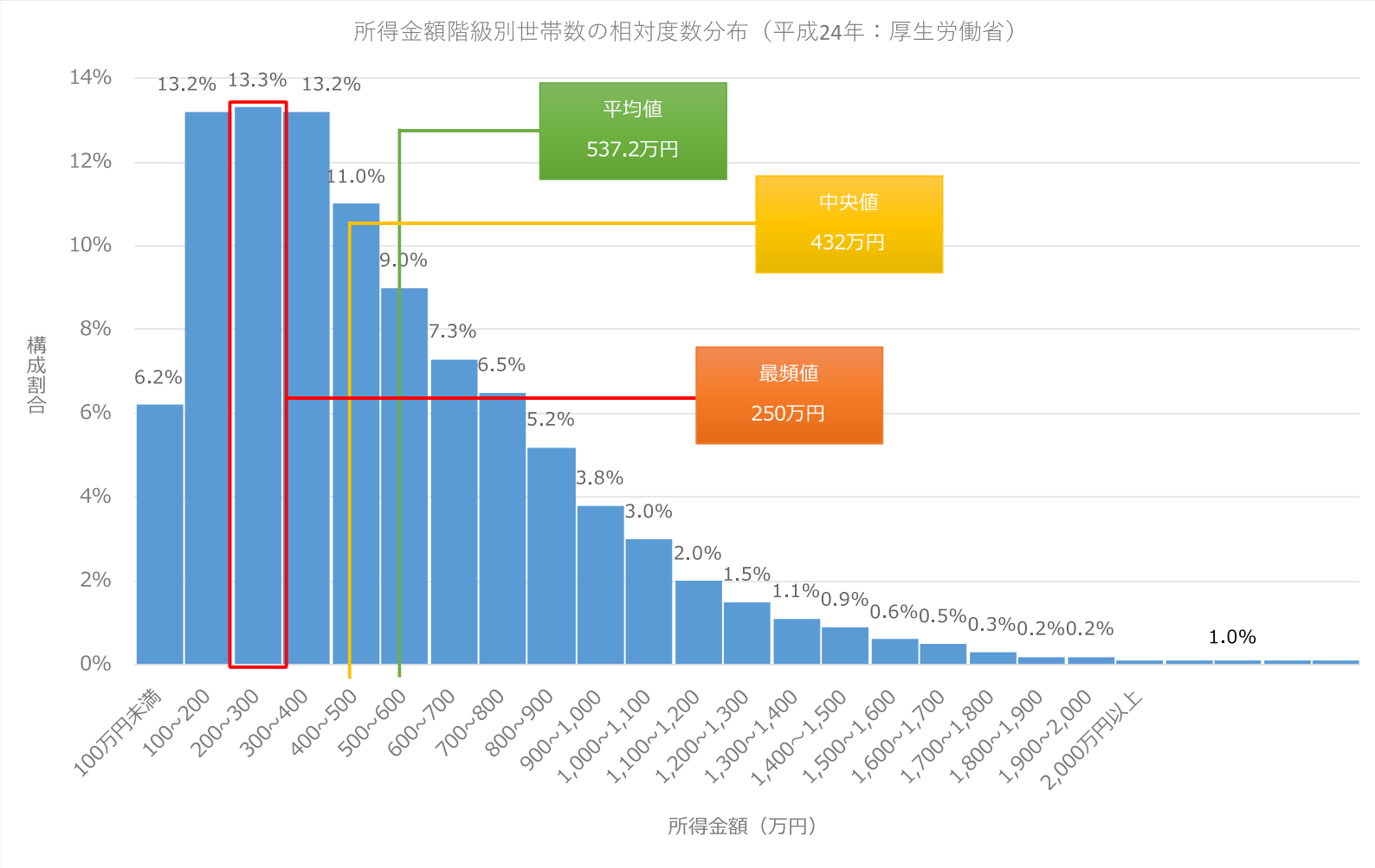
- 代表値の名称と特徴を把握する。

	平均値	中央値	最頻値
定義	合計をサンプル数で割ったもの	順番に並べて真ん中に来る値	最も出現頻度が高い値
メリット	わかりやすい	歪み（外れ値）に強い	ボリュームゾーンが分かる
デメリット	歪みに弱い	計算が面倒	全体傾向がつかめない

- ビジネスデータは「平均値 ≠ 中央値」であることが多いことに注意する。

代表値歪みの典型例

- 代表値が歪む典型例は「年収分布」。平均値・中央値・最頻値が大きく異なる。

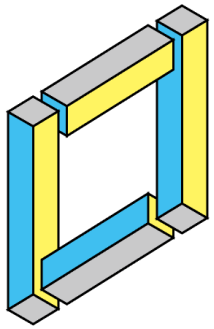


データを見るべきポイント

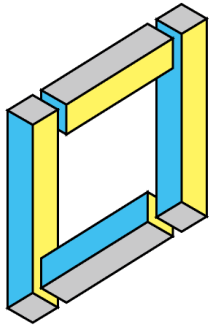
項目	内容とポイント
データ数	結果がどの程度のデータ数から得られたかを把握する。 仮説検定を行う場合は最低でも30データはほしい。
平均値	まずは平均値を見ることからスタートする。
標準偏差	標準偏差はデータのブレの範囲を表す。ブレが大きいか小さいかを把握する。
最小値	最小値と最大値はデータの信頼性を見る上で重要なポイント。平均値の次に見るべき数字。
25%点	25%点と75%点は偏りが前半にあるか後半にあるかを見る上で重要。
中央値（50%点）	中央値を見ることで、データの偏りを把握する事ができる。平均値が実感と異なる場合は中央値を探る。
75%点	25%点と75%点は偏りが前半にあるか後半にあるかを見る上で重要。
最大値	最小値と最大値はデータの信頼性を見る上で重要なポイント。平均値の次に見るべき数字。
最頻値	データのピークがどのあたりにあるのかを押さえる。データが多い場合は最頻値の上位3つ程押さえておくとよい。

まとめ

- データ分析は様々な領域から構成される。自身が学習している範囲を理解しよう。
- 分析もビジネスコミュニケーションの一種である。
- データには大きな分類として、「質的変数」と「量的変数」がある。
- 代表値は「平均値」「中央値」「最頻値」がある。
  それぞれのメリット・デメリットを理解して使い分ける。
- データを入手したら、見るべきポイントに従ってデータの全体像を把握しよう。

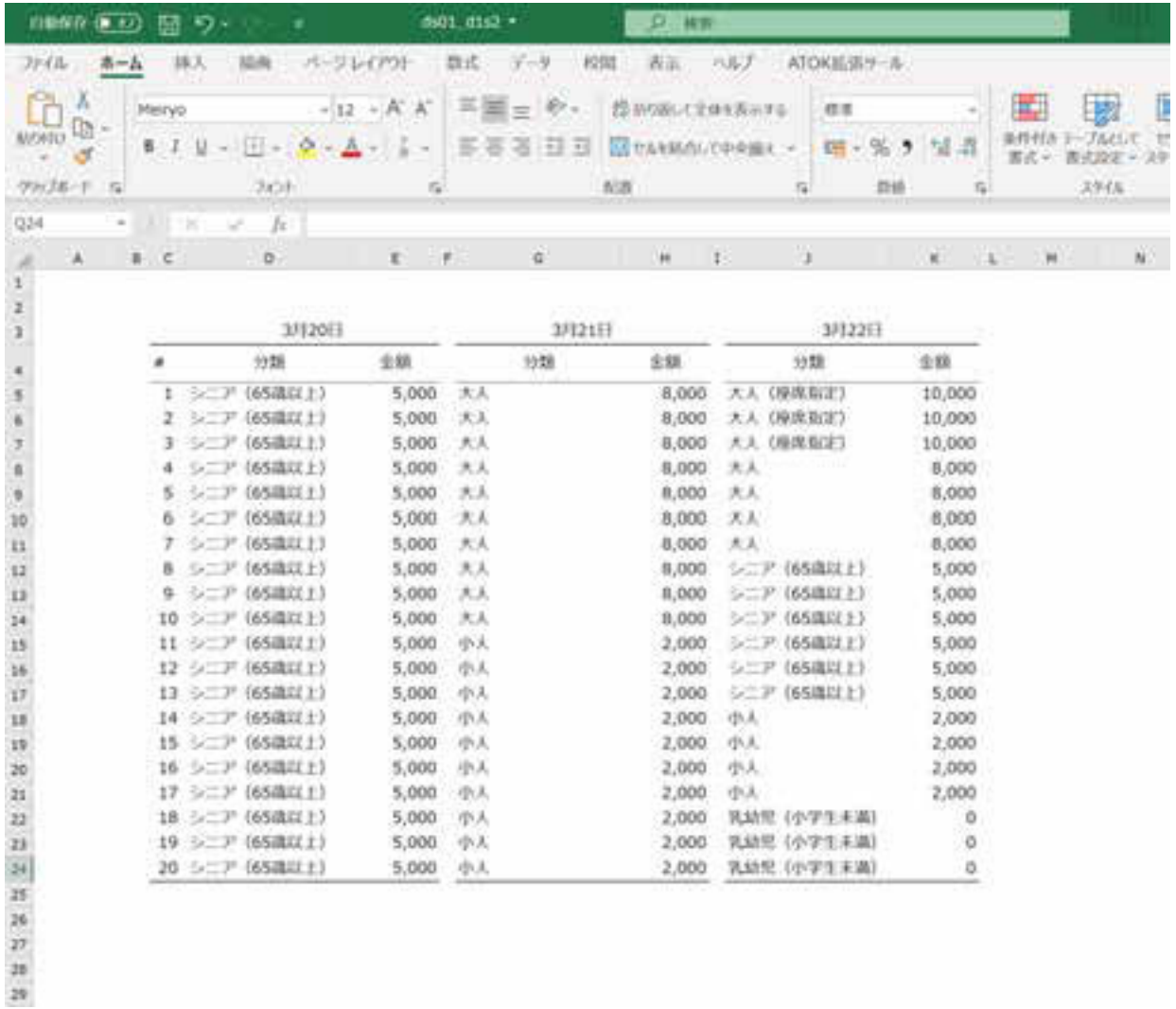


ハンズオン



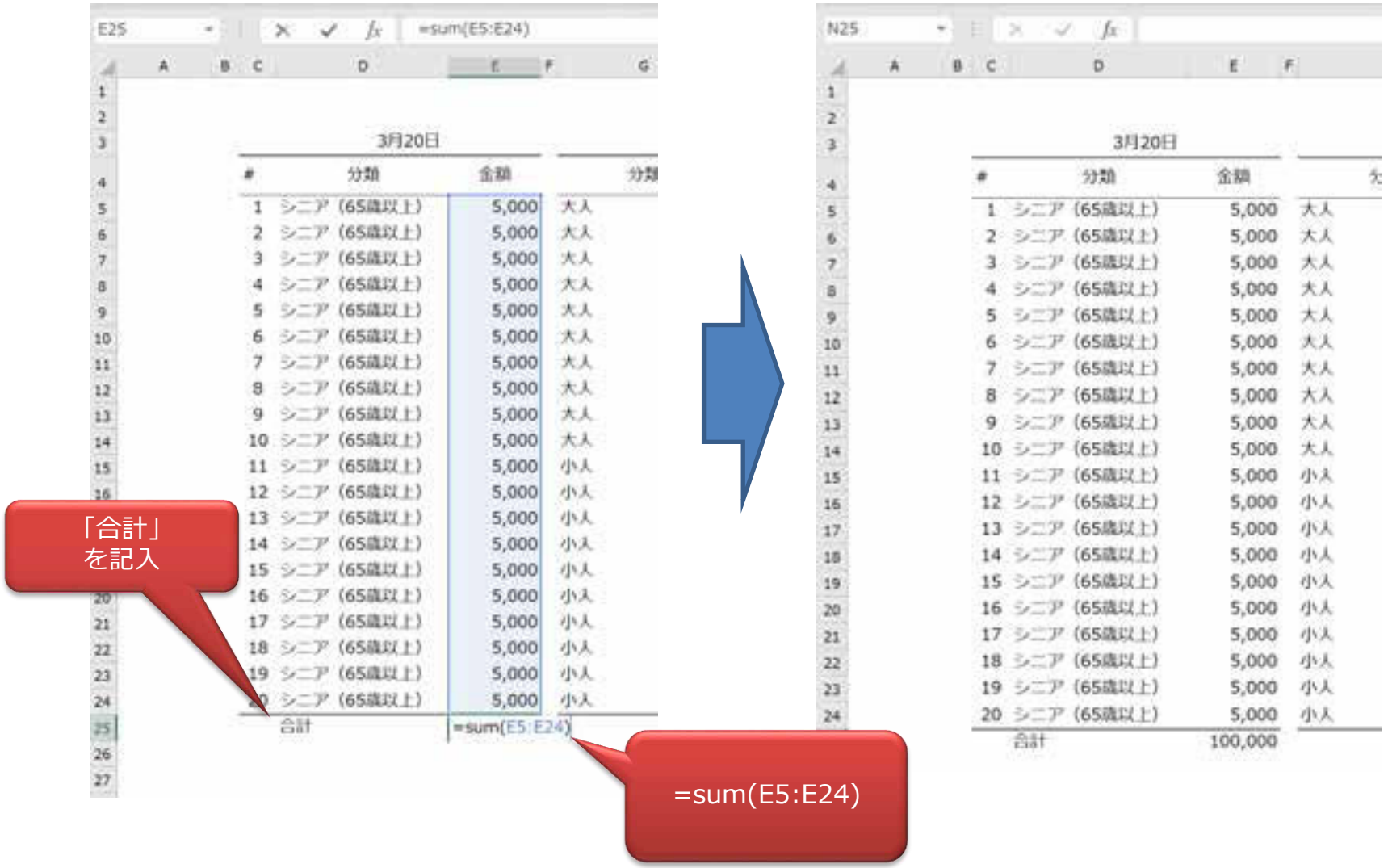
ファイルを開く

- ファイル「演習データD1_1.xlsx」を開きましょう。

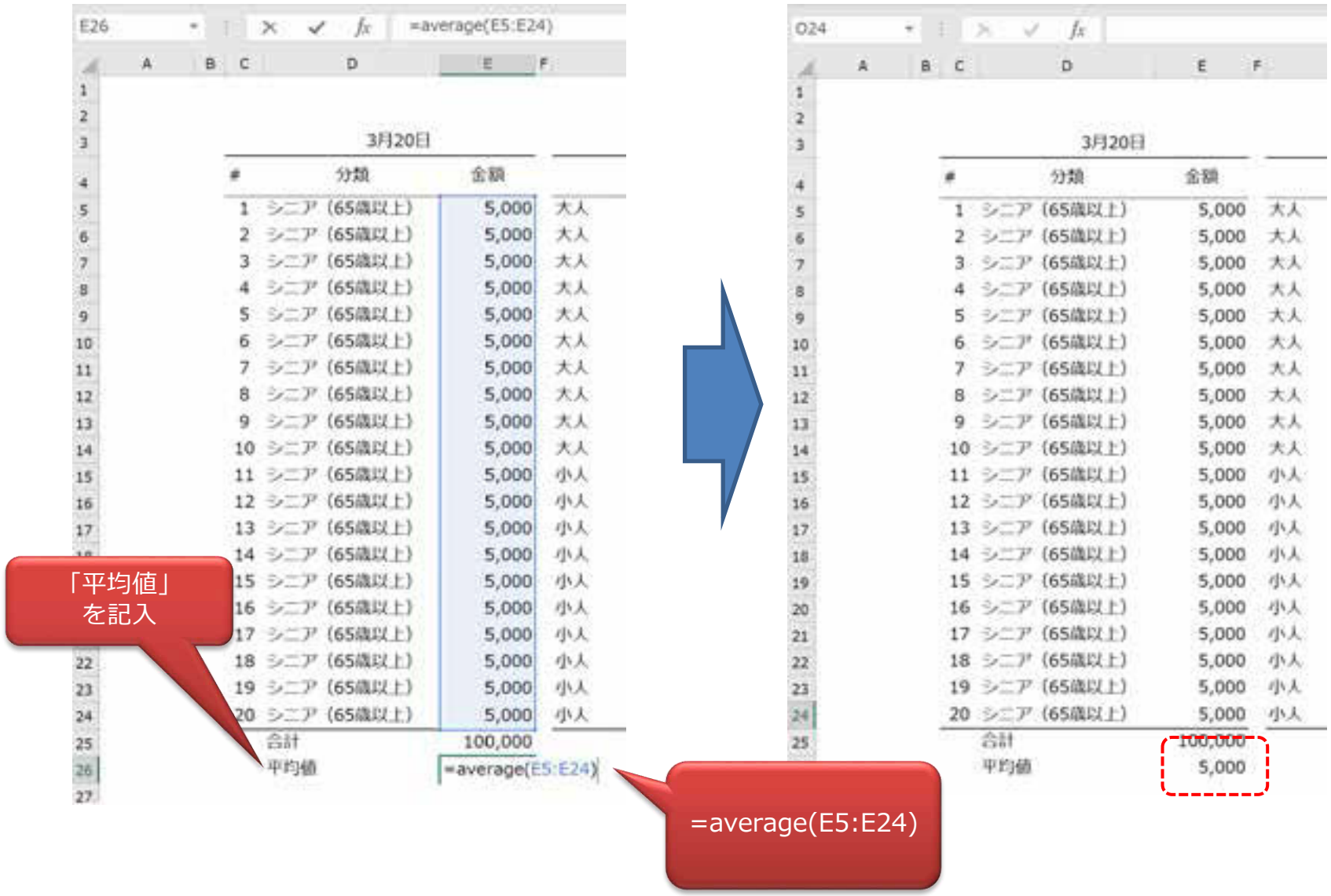


	3月20日		3月21日		3月22日	
#	分類	金額	分類	金額	分類	金額
1	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
2	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
3	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
4	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
5	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
6	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
7	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
8	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
9	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
10	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
11	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
12	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
13	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
14	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
15	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
16	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
17	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
18	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
19	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
20	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0

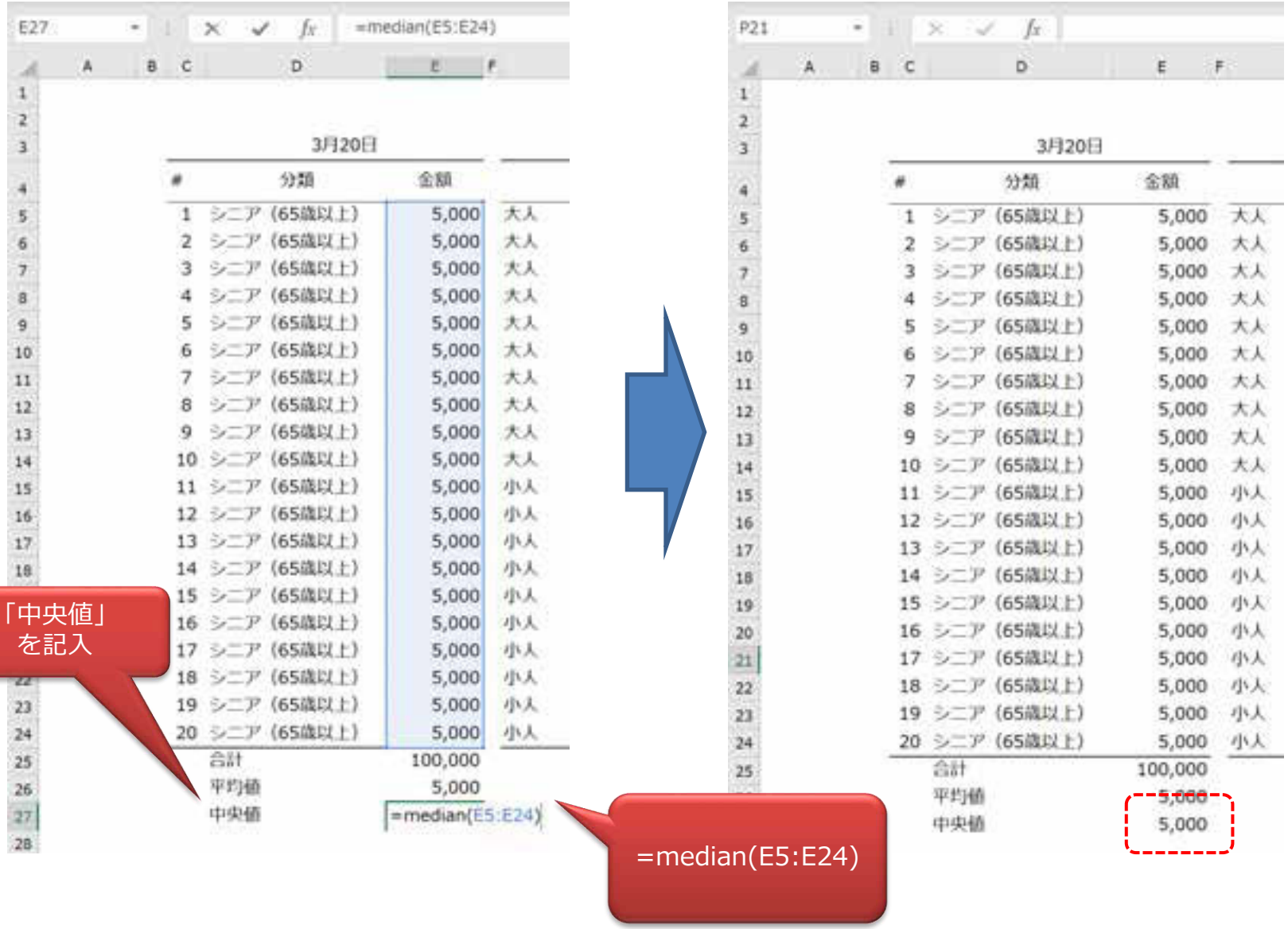
合計を求める


E25
=sum(E5:E24)
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 =sum(E5:E24)
「合計」
を記入
=sum(E5:E24)
N25
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000

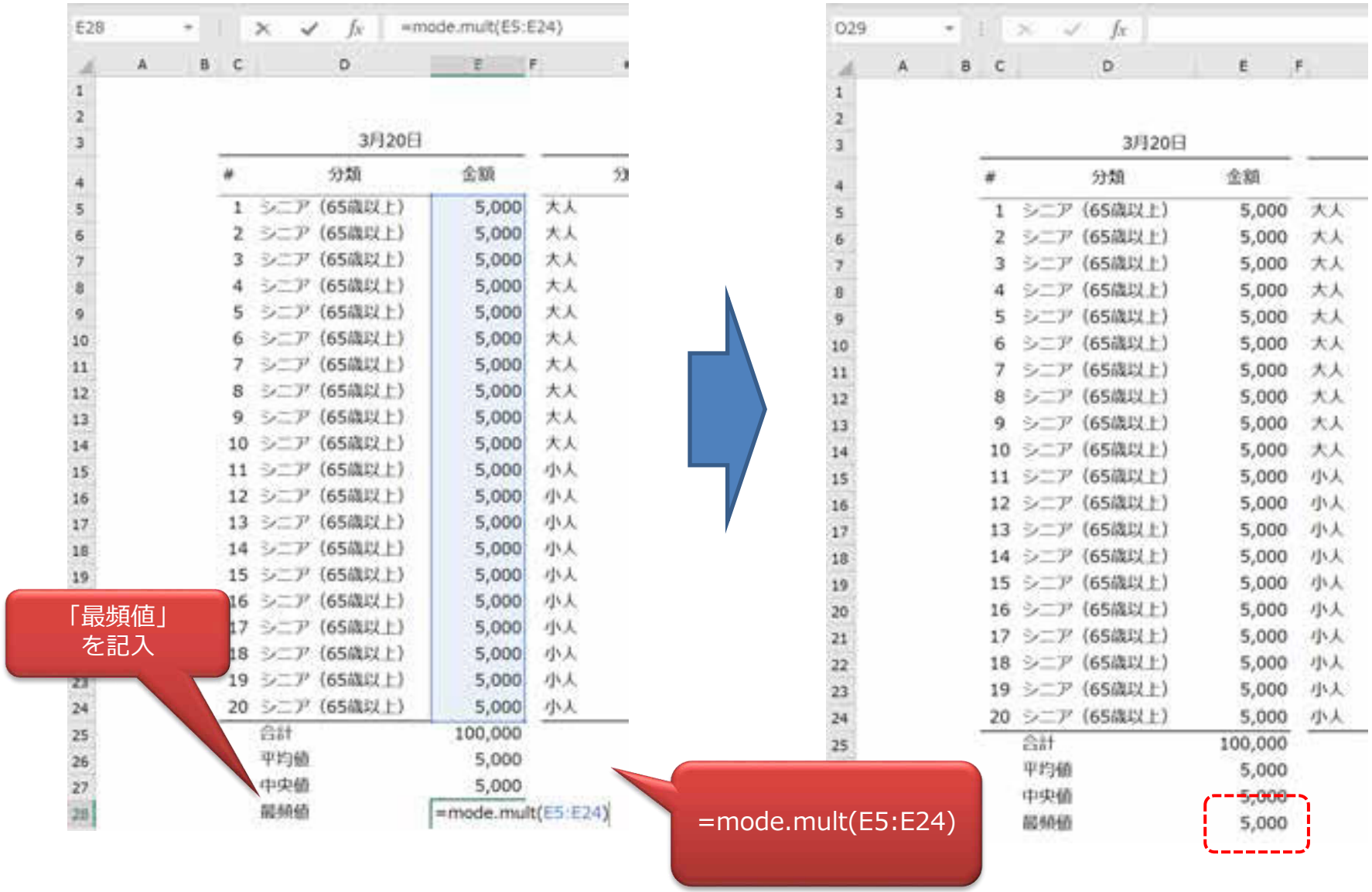
平均値を求める


E26
=average(E5:E24)
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000
平均値 =average(E5:E24)
「平均値」を記入
O24
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000
平均値 5,000
=average(E5:E24)

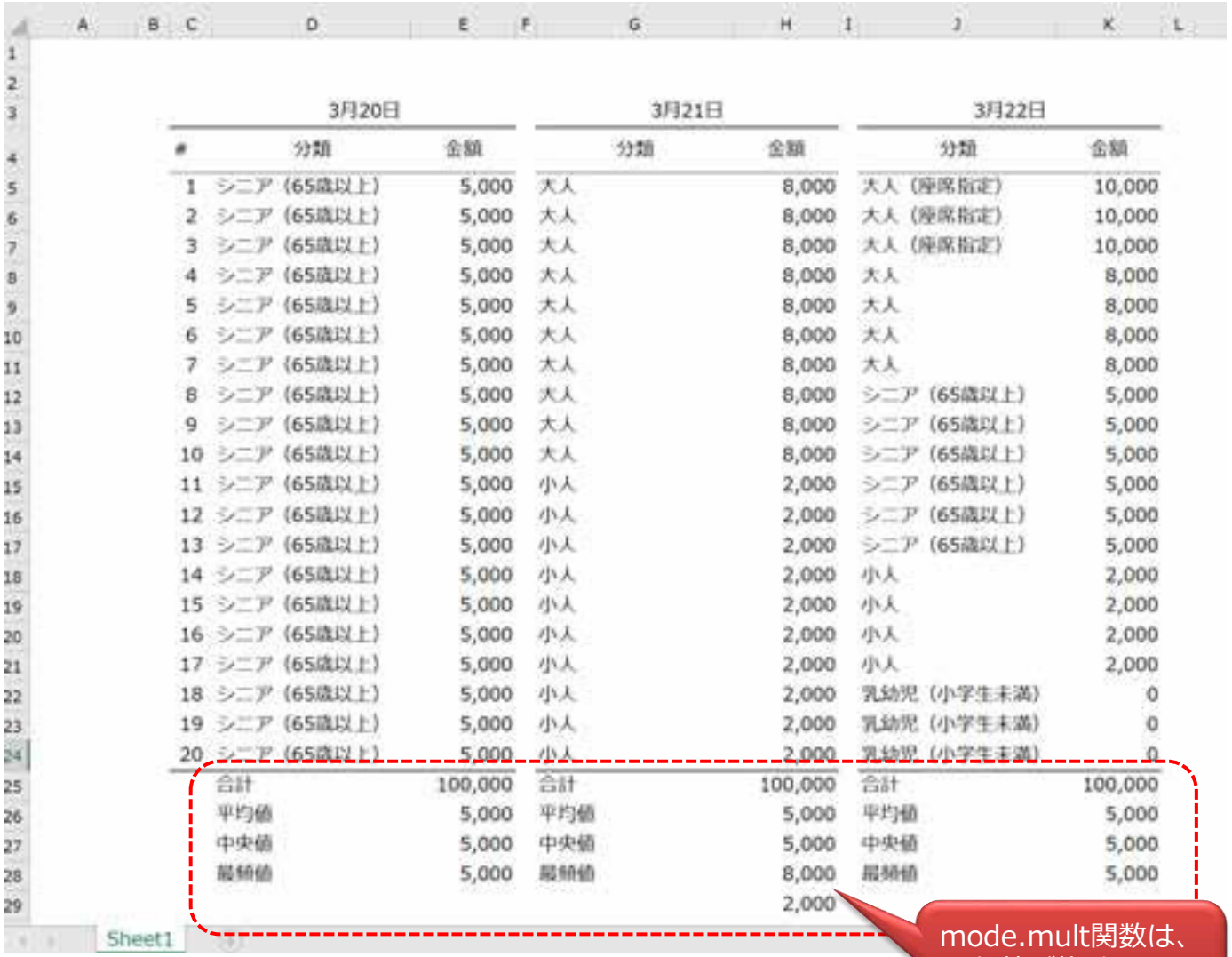
中央値を求める


E27
=median(E5:E24)
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000
平均値 5,000
中央値 =median(E5:E24)
「中央値」を記入
=median(E5:E24)
P21
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000
平均値 5,000
中央値 5,000

最頻値を求める


E28
=mode.mult(E5:E24)
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000
平均値 5,000
中央値 5,000
最頻値 =mode.mult(E5:E24)
「最頻値」を記入
O29
3月20日
分類 金額
1 シニア（65歳以上） 5,000 大人
2 シニア（65歳以上） 5,000 大人
3 シニア（65歳以上） 5,000 大人
4 シニア（65歳以上） 5,000 大人
5 シニア（65歳以上） 5,000 大人
6 シニア（65歳以上） 5,000 大人
7 シニア（65歳以上） 5,000 大人
8 シニア（65歳以上） 5,000 大人
9 シニア（65歳以上） 5,000 大人
10 シニア（65歳以上） 5,000 大人
11 シニア（65歳以上） 5,000 小人
12 シニア（65歳以上） 5,000 小人
13 シニア（65歳以上） 5,000 小人
14 シニア（65歳以上） 5,000 小人
15 シニア（65歳以上） 5,000 小人
16 シニア（65歳以上） 5,000 小人
17 シニア（65歳以上） 5,000 小人
18 シニア（65歳以上） 5,000 小人
19 シニア（65歳以上） 5,000 小人
20 シニア（65歳以上） 5,000 小人
合計 100,000
平均値 5,000
中央値 5,000
最頻値 5,000
=mode.mult(E5:E24)

3パターンで実行



#	3月20日 分類	3月20日 金額	3月21日 分類	3月21日 金額	3月22日 分類	3月22日 金額
1	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
2	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
3	シニア（65歳以上）	5,000	大人	8,000	大人（座席指定）	10,000
4	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
5	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
6	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
7	シニア（65歳以上）	5,000	大人	8,000	大人	8,000
8	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
9	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
10	シニア（65歳以上）	5,000	大人	8,000	シニア（65歳以上）	5,000
11	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
12	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
13	シニア（65歳以上）	5,000	小人	2,000	シニア（65歳以上）	5,000
14	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
15	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
16	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
17	シニア（65歳以上）	5,000	小人	2,000	小人	2,000
18	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
19	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
20	シニア（65歳以上）	5,000	小人	2,000	乳幼児（小学生未満）	0
	合計	100,000	合計	100,000	合計	100,000
	平均値	5,000	平均値	5,000	平均値	5,000
	中央値	5,000	中央値	5,000	中央値	5,000
	最頻値	5,000	最頻値	8,000	最頻値	5,000
				2,000		

Sheet1

mode.mult関数は、最頻値が複数ある場合も出力できます。

統計学の基本

統計学の基礎　④代表値の読み取りと意義

ワーク

データを眺めて全体を把握

データを眺めて、歪みを予想してください。

ファイル「演習データE1.xlsx」を開きましょう。

データ全体を”眺めて”ください。

どのような印象をもったか、グループで意見交換しましょう。

その後、実際に代表値を算出し、何が読み取れるか考えて
みましょう。

ワークの進め方（１）



後で見て分かるように
テーマを書いておきましょう

データを眺めた印象を
グループで話し合いメモしま
しょう。

以下のようなことを予想してみましょう。

代表値は、それぞれどのあたりでしょうか？
代表値に歪みは、ありそうでしょうか？

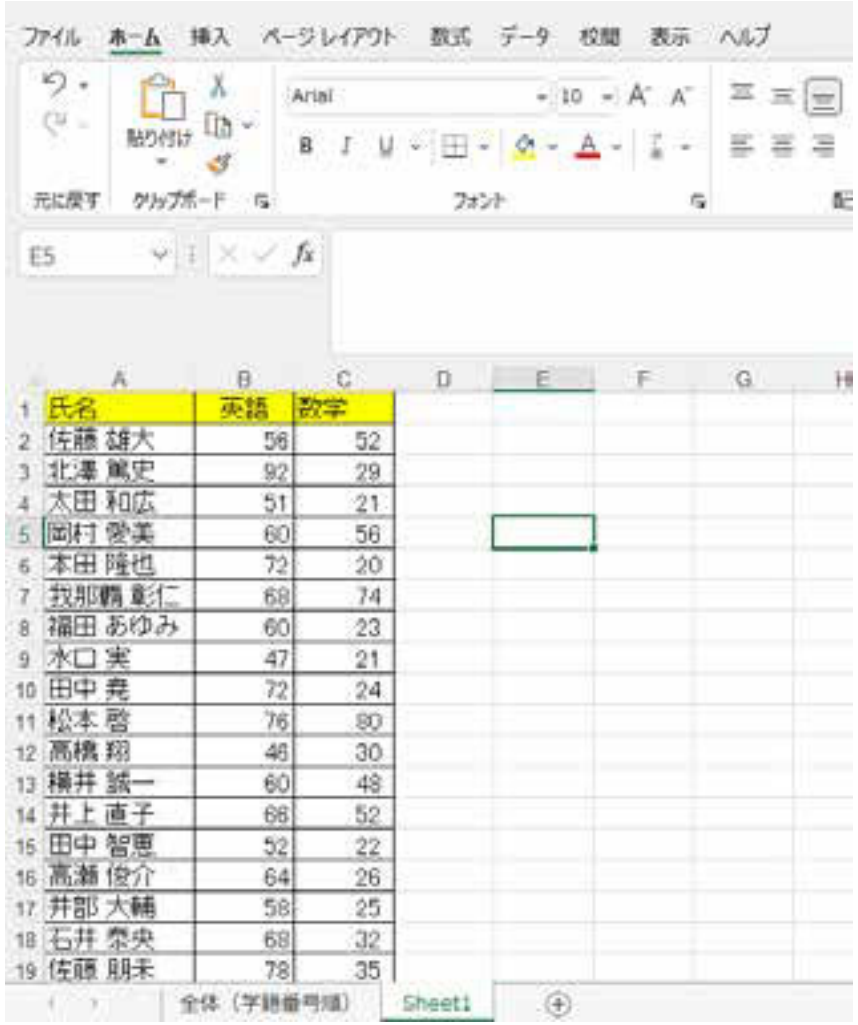
ワークの進め方（２）

では、実際に代表値を算出してみましょう。

ハンズオン（答えあわせ）

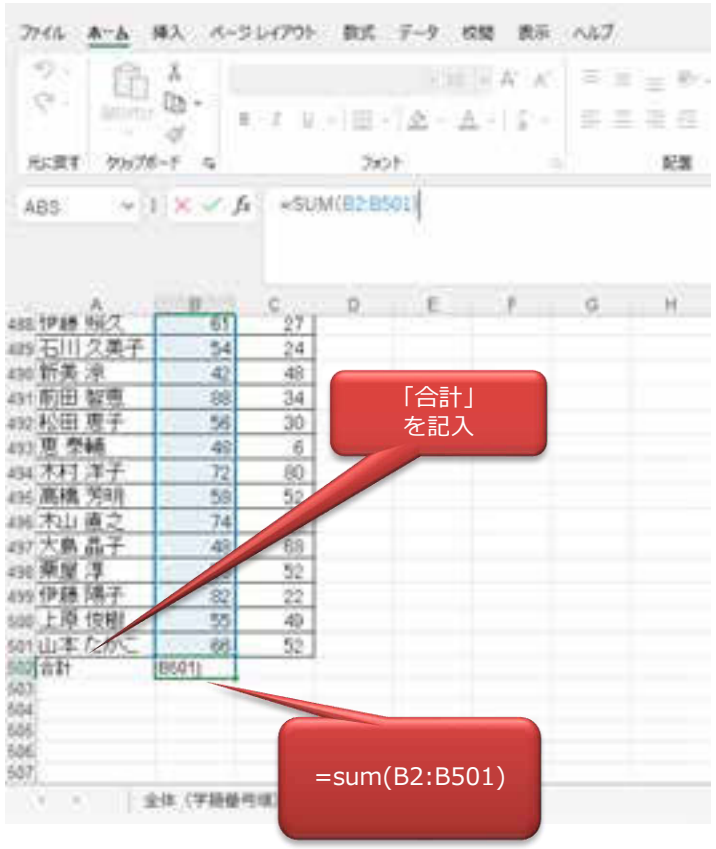
ファイルを開く

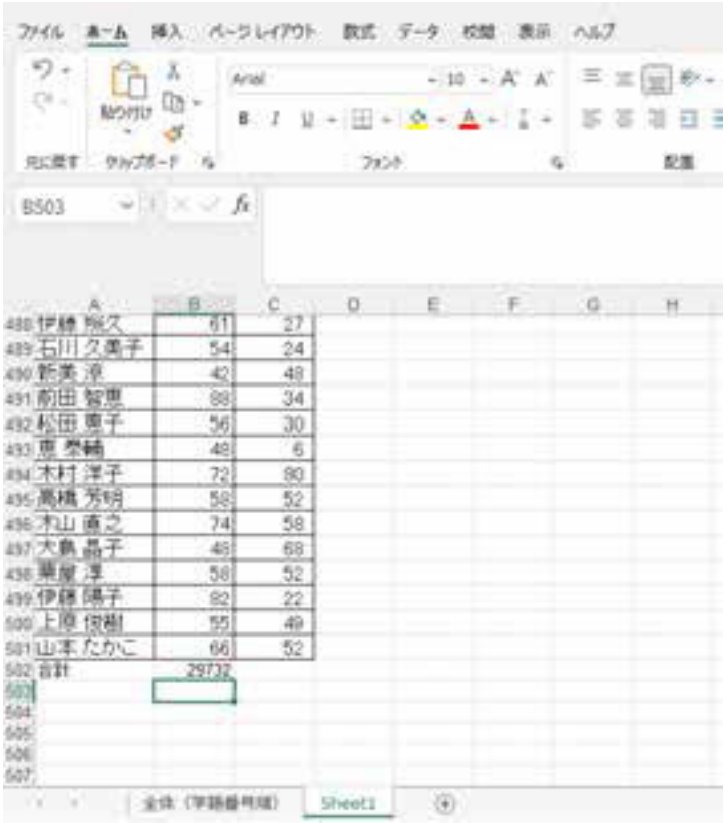
- ファイル「演習データE1.xlsx」を開きましょう。



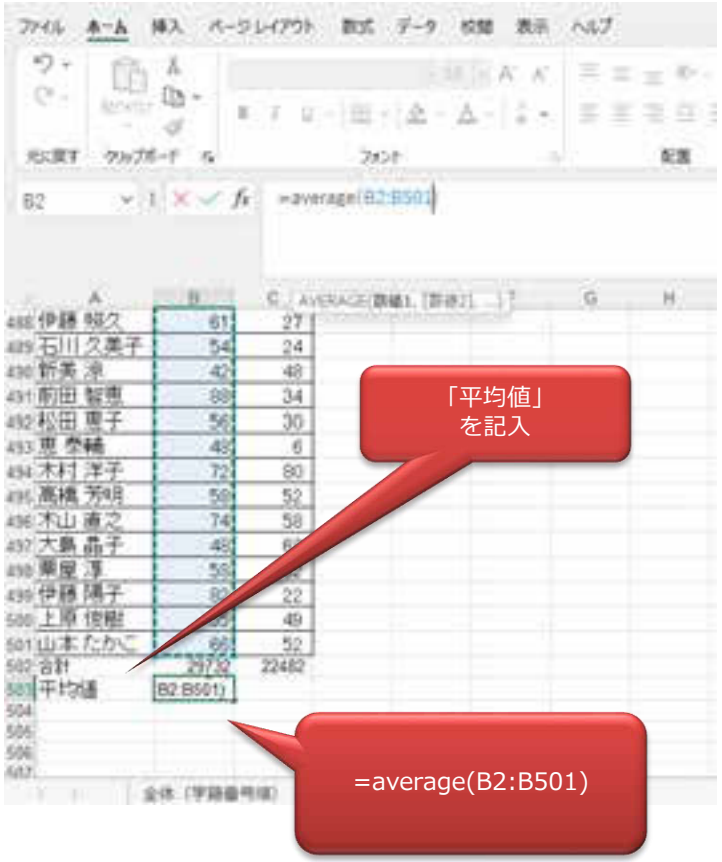
	A	B	C
1	氏名	英語	数学
2	佐藤 雄大	56	52
3	北澤 篤史	92	29
4	太田 和広	51	21
5	岡村 愛美	60	56
6	本田 隆也	72	20
7	我那覇 彰仁	68	74
8	福田 あゆみ	60	23
9	水口 実	47	21
10	田中 亮	72	24
11	松本 啓	76	80
12	高橋 翔	46	30
13	横井 誠一	60	48
14	井上 直子	68	52
15	田中 智恵	52	22
16	高瀬 俊介	64	26
17	井部 大輔	58	25
18	石井 泰央	68	32
19	佐藤 朋未	78	35

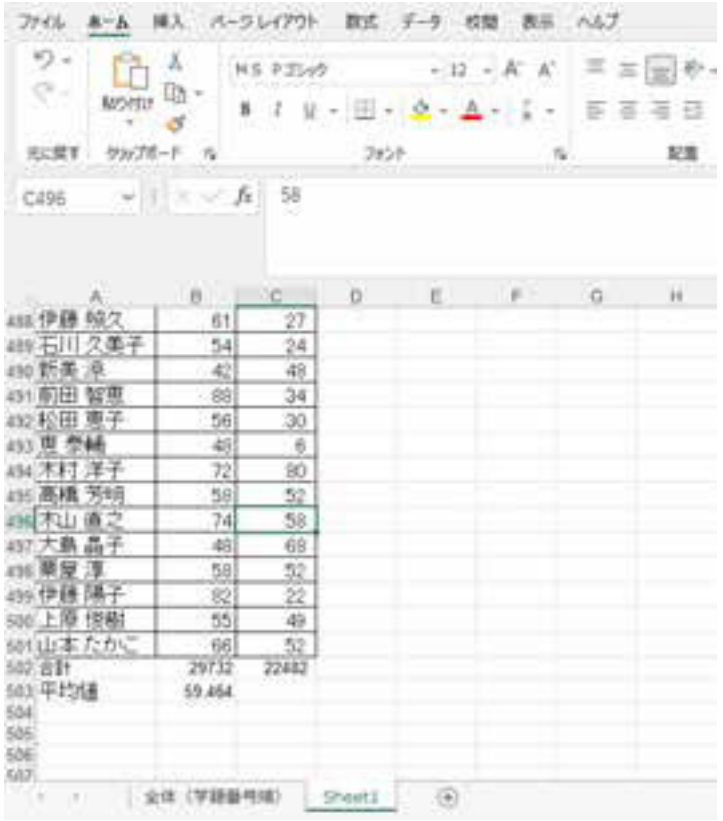
合計を求める



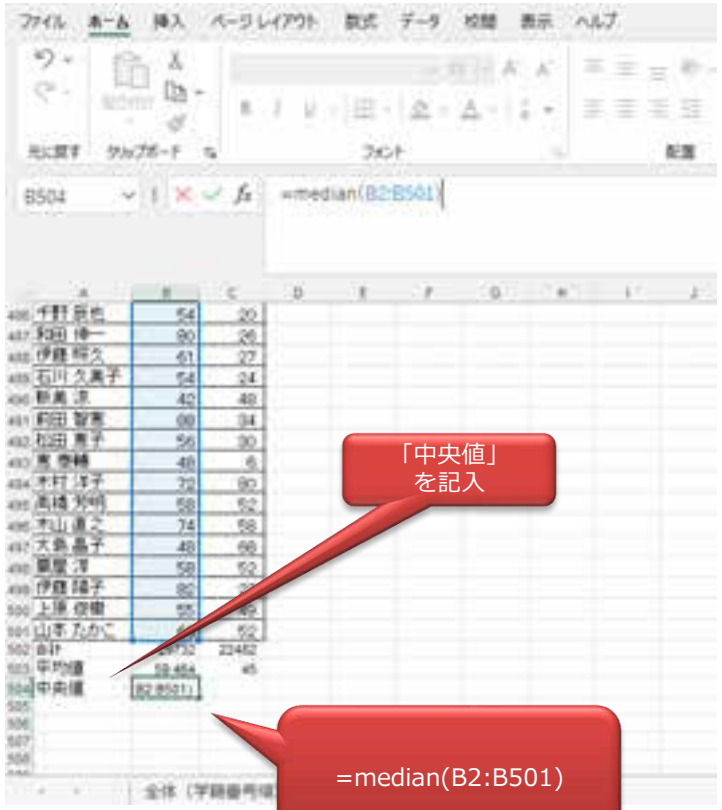


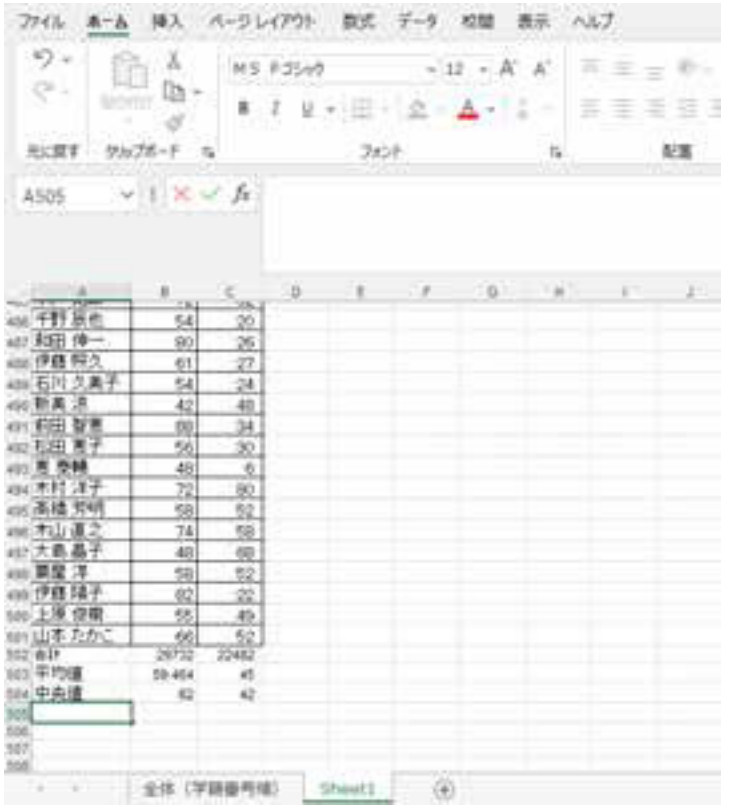
平均値を求める





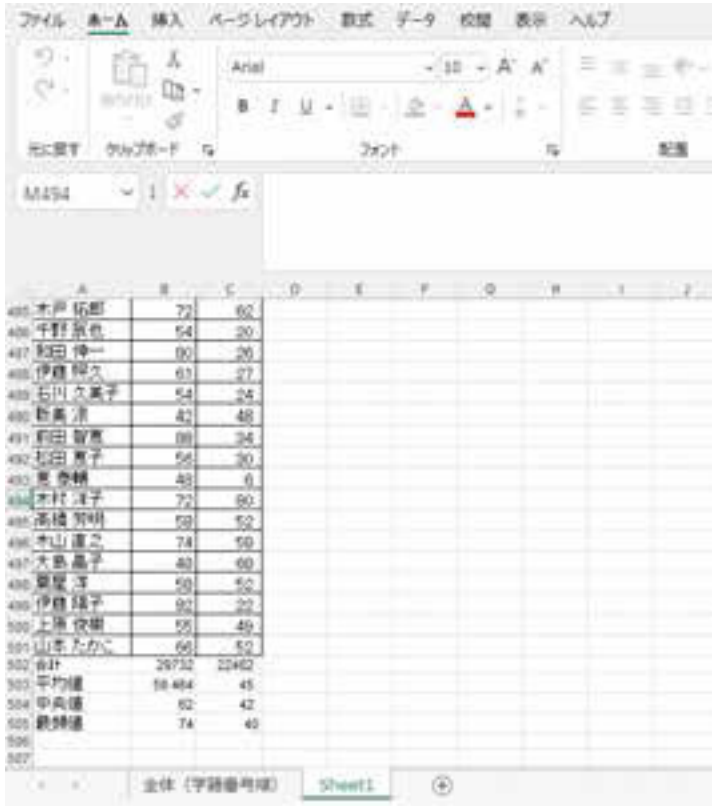
中央値を求める





最頻値を求める





2パターンで実行



	A	B	C	D
490	新美 涼	42	48	
491	前田 智恵	88	34	
492	松田 恵子	56	30	
493	恵 泰輔	48	6	
494	木村 洋子	72	80	
495	高橋 芳明	58	52	
496	木山 直之	74	58	
497	大島 晶子	48	68	
498	粟屋 淳	58	52	
499	伊藤 陽子	82	22	
500	上原 俊樹	55	49	
501	山本 たかこ	66	52	
502	合計	29732	22482	
503	平均値	59.464	45	
504	中央値	62	42	
505	最頻値	74	40	
506				

mode.mult関数は、最頻値が
複数ある場合も出力できます。

英語

平均値：59.5点

中央値：62点

最頻値：74点

数学

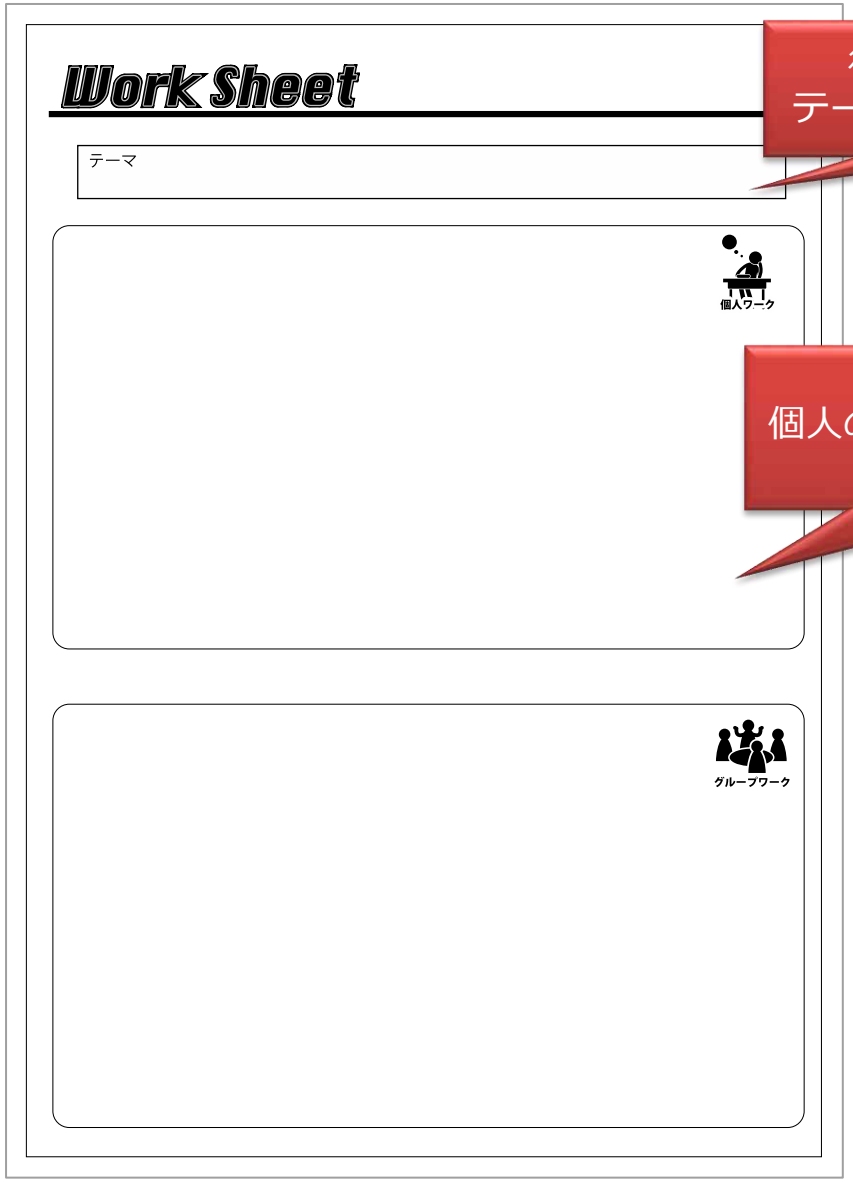
平均値：45点

中央値：42点

最頻値：40点

ワークの続き

ワークの進め方（２）

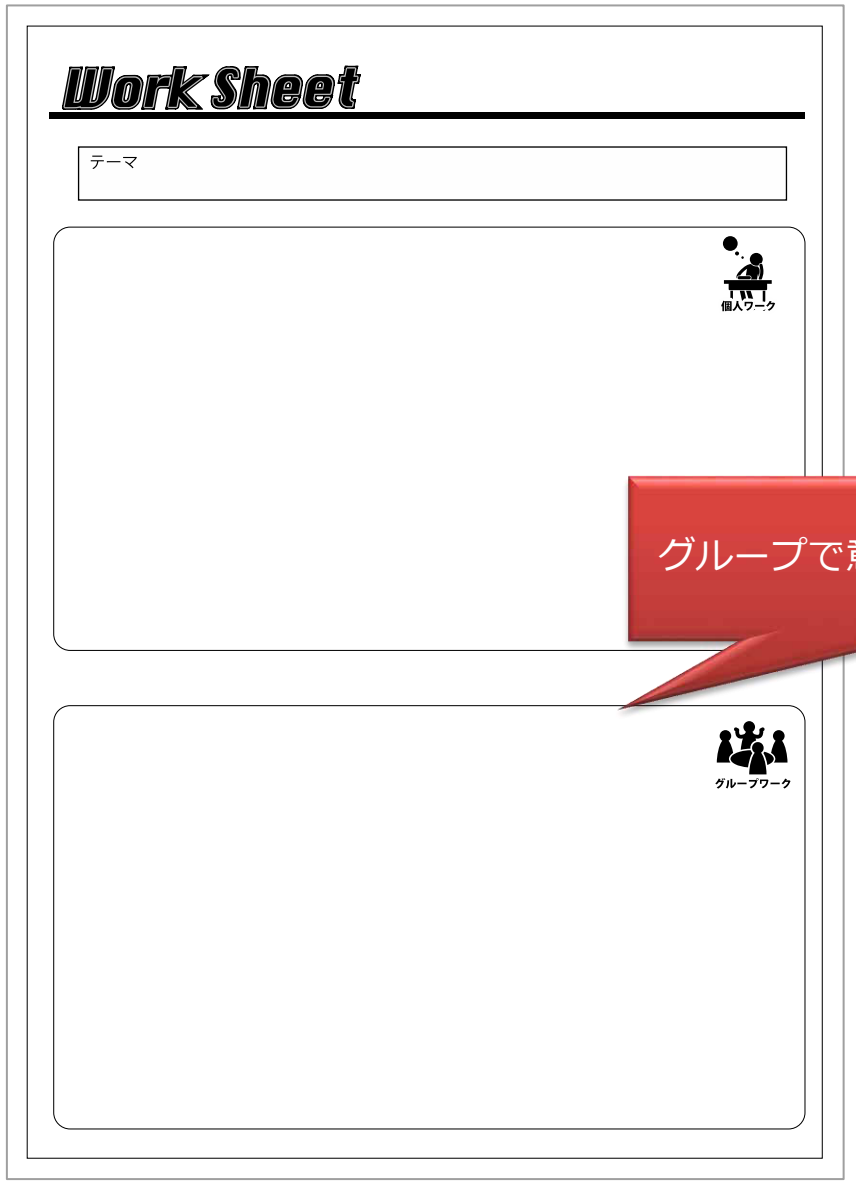


後で見て分かるように
テーマを書いておきましょう

個人の考えを書いておきましょう

算出した代表値から
どのようなことが読み取れますか？
個人の考えを整理しましょう。

ワークの進め方（３）



グループで意見交換をしましょう。

個人で読み取ったことを
グループ内で意見交換しましょう。
結論は、出さなくても良いですが
「全員が納得する読み取り」はないか
議論して探ってみましょう。

ワークの進め方（４）

グループごとの意見を全体に共有しましょう。

グループで出た意見の中で
良かった・興味深かった
変わっている（たぶん他では出てないだろう）
自分が気に入っている（元々自分の意見でもOK）
特に理由はないけど、皆に聞いて欲しい

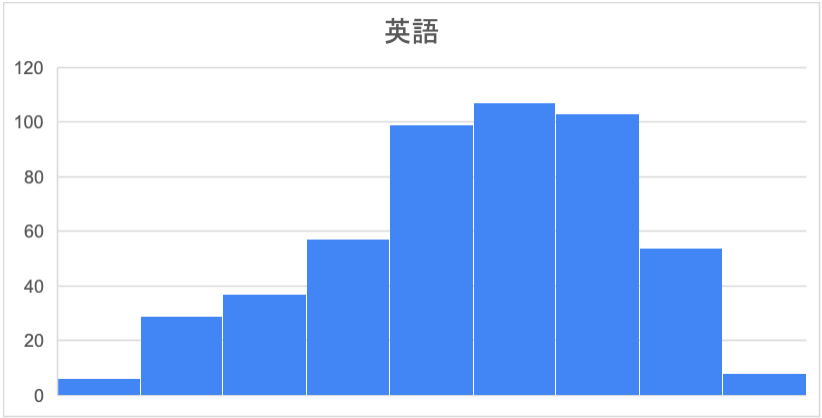
というものを、共有してください。

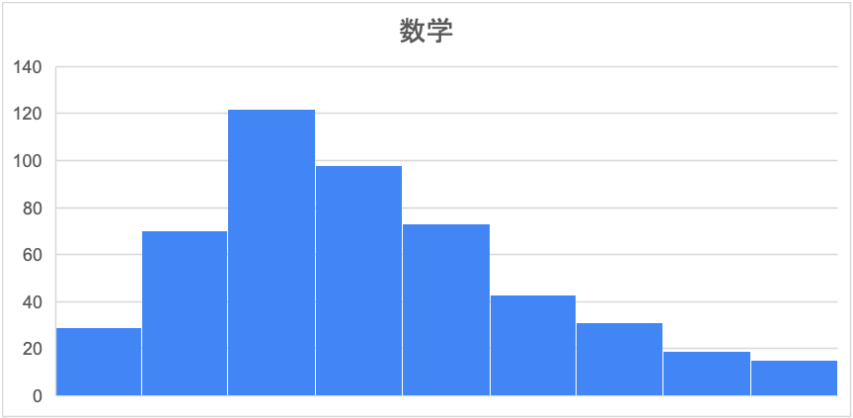
※他のグループの説明から新しい視点を見つけてください。



代表値から見えること

二つのデータの分布の外観は、下記のようになります。





3つの代表値からだけでも、例えば

「数学より英語の方が、高得点に偏っている」

ということくらいは分かります。

しかし、代表値からだけでは分からないことも多いので

他にもいろいろな統計値が必要になってくるのです。

統計学の基本

統計学の基礎　⑤散布度と算出

散布度とは？

- 散布度とは、平均値の周りの「データのばらつきの程度」を表します。
- 代表値に加えて、散布度が使えるようになるとデータの見方がより豊かになります。

偏 差

- 偏差とは、「平均値からのズレ」を表します。
  - 平均値からのズレが大きいほど数字が大きくなります。
  - 平均値よりも大きい場合は正の値、平均値よりも小さい場合は負の値を取ります。

偏 差

3月22日

#	分類	金額	偏差
1	大人（座席指定）	10,000	5,000
2	大人（座席指定）	10,000	5,000
3	大人（座席指定）	10,000	5,000
4	大人	8,000	3,000
5	大人	8,000	3,000
6	大人	8,000	3,000
7	大人	8,000	3,000
8	シニア（65歳以上）	5,000	0
9	シニア（65歳以上）	5,000	0
10	シニア（65歳以上）	5,000	0
11	シニア（65歳以上）	5,000	0
12	シニア（65歳以上）	5,000	0
13	シニア（65歳以上）	5,000	0
14	小人	2,000	-3,000
15	小人	2,000	-3,000
16	小人	2,000	-3,000
17	小人	2,000	-3,000
18	乳幼児（小学生未満）	0	-5,000
19	乳幼児（小学生未満）	0	-5,000
20	乳幼児（小学生未満）	0	-5,000
	合計	100,000	
	平均値	5,000	

偏差は「金額ー平均値」で求まる。
平均値より大きいとプラス、平均値より小さいとマイナスの値をとる

平均値

平均偏差

- 平均偏差とは、「偏差の絶対値の平均値」のことで、
  平均値からのばらつきの大きさを表します。
  - 全体的にデータがどの程度ばらついているのかを調べるために用います。
    - 偏差は平均値からのズレであるため、「偏差の合計は必ずゼロ」となります。
    - このため、偏差の絶対値の平均値をとることで、「平均値からのズレの大きさの平均値」を求めることができます。
  - 値が大きいほど平均値からのばらつきが大きくなります。
  - 全部の値が同じ場合、平均偏差はゼロになります。

平均偏差

3月22日

#	分類	金額	偏差	|偏差|
1	大人（座席指定）	10,000	5,000	5,000
2	大人（座席指定）	10,000	5,000	5,000
3	大人（座席指定）	10,000	5,000	5,000
4	大人	8,000	3,000	3,000
5	大人	8,000	3,000	3,000
6	大人	8,000	3,000	3,000
7	大人	8,000	3,000	3,000
8	シニア（65歳以上）	5,000	0	0
9	シニア（65歳以上）	5,000	0	0
10	シニア（65歳以上）	5,000	0	0
11	シニア（65歳以上）	5,000	0	0
12	シニア（65歳以上）	5,000	0	0
13	シニア（65歳以上）	5,000	0	0
14	小人	2,000	-3,000	3,000
			-3,000	3,000
			-3,000	3,000
			-3,000	3,000
18	乳幼児（小学生未満）		-5,000	5,000
19	乳幼児（小学生未満）		-5,000	5,000
20	乳幼児（小学生未満）	0	-5,000	5,000
	合計	100,000	0	54,000
	平均値	5,000	0	2,700

偏差の平均値がゼロになる原因は符号の正負によるもの。
符号の影響をなくすため、偏差の絶対値を計算する。

偏差の平均値は必ずゼロになる。

平均偏差: 2,700円
金額は平均値5,000円に対して
±2,700円程度でばらついている

分 散

- 分散とは、「偏差の二乗の平均値」のことです。
  - 平均偏差では、偏差の大きさを求めるために偏差の絶対値を用いましたが、もう一つの方法として、偏差の二乗を用いるというアプローチがあります。
  - 本講座では標準偏差につながるステップとしての計算方法のみ紹介し、詳しくは触れません。

分 散

3月22日

#	分類	金額	偏差	|偏差|	(偏差)^2
1	大人（座席指定）	10,000	5,000	5,000	25,000,000
2	大人（座席指定）	10,000	5,000	5,000	25,000,000
3	大人（座席指定）	10,000	5,000	5,000	25,000,000
4	大人	8,000	3,000	3,000	9,000,000
5	大人	8,000	3,000	3,000	9,000,000
6	大人	8,000	3,000	3,000	9,000,000
7	大人	8,000	3,000	3,000	9,000,000
8	シニア（65歳以上）	5,000	0	0	0
9	シニア（65歳以上）	5,000	0	0	0
10	シニア（65歳以上）	5,000	0	0	0
11	シニア（65歳以上）	5,000	0	0	0
12	シニア（65歳以上）	5,000	0	0	0
13	シニア（65歳以上）	5,000	0	0	0
14	小人	2,000	-3,000	3,000	9,000,000
15	小人	2,000	-3,000	3,000	9,000,000
16	小人	2,000	-3,000	3,000	9,000,000
17	小人	2,000	-3,000	3,000	9,000,000
18	乳幼児（小学生未満）	0	-5,000	5,000	25,000,000
19	乳幼児（小学生未満）	0	-5,000	5,000	25,000,000
20	乳幼児（小学生未満）	0	-5,000	5,000	25,000,000
	合計	100,000	0	54,000	222,000,000
	平均値	5,000	0	2,700	11,100,000

偏差の平均値がゼロになる原因は符号の正負によるもの。
符号の影響をなくすため、偏差の二乗を計算する。

分散: 11,100,000円
二乗した量なので、非常に大きくなってしまう。

標準偏差

- 標準偏差とは、
  「偏差の二乗の平均値の平方根=分散の平方根」のことで、
  平均偏差同様に平均値からのばらつきの大きさを表します。
  - 全体的にデータがどの程度ばらついているのかを調べるために用います。
    - 数学的に利用しやすいことから、平均値からのばらつきの大きさは標準偏差が頻繁に用いられます。
  - 値が大きいほど平均値からのばらつきが大きくなります。
  - 全部の値が同じ場合、平均偏差はゼロになります。

標準偏差

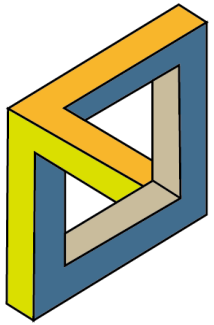
3月22日

#	分類	金額	偏差	|偏差|	(偏差)^2
1	大人（座席指定）	10,000	5,000	5,000	25,000,000
2	大人（座席指定）	10,000	5,000	5,000	25,000,000
3	大人（座席指定）	10,000	5,000	5,000	25,000,000
4	大人	8,000	3,000	3,000	9,000,000
5	大人	8,000	3,000	3,000	9,000,000
6	大人	8,000	3,000	3,000	9,000,000
7	大人	8,000	3,000	3,000	9,000,000
8	シニア（65歳以上）	5,000	0	0	0
9	シニア（65歳以上）	5,000	0	0	0
10	シニア（65歳以上）	5,000	0	0	0
11	シニア（65歳以上）	5,000	0	0	0
12	シニア（65歳以上）	5,000	0	0	0
13	シニア（65歳以上）	5,000	0	0	0
14	小人	2,000	-3,000	3,000	9,000,000
15	小人	2,000	-3,000	3,000	9,000,000
16	小人	2,000	-3,000	3,000	9,000,000
17	小人	2,000	-3,000	3,000	9,000,000
18	乳幼児（小学生未満）	0	-5,000	5,000	25,000,000
19	乳幼児（小学生未満）	0	-5,000	5,000	25,000,000
20	乳幼児（小学生未満）	0	-5,000	5,000	25,000,000
	合計	100,000	0	54,000	222,000,000
	平均値	5,000	0	2,700	11,100,000
	平方根				3,331.7

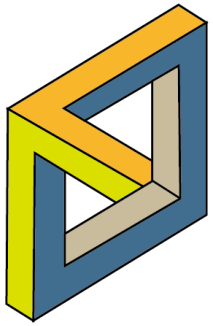
標準偏差: 3,331.7円
二乗した分を戻すため、分散の平方根を求める。

標準偏差のイメージ

- 標準偏差は「ダイバーシティ」
  - 全データが同じだと0になる。
  - データのばらつきが大きいほど（多様性が高いほど）標準偏差の値は大きくなる。



ハンズオン

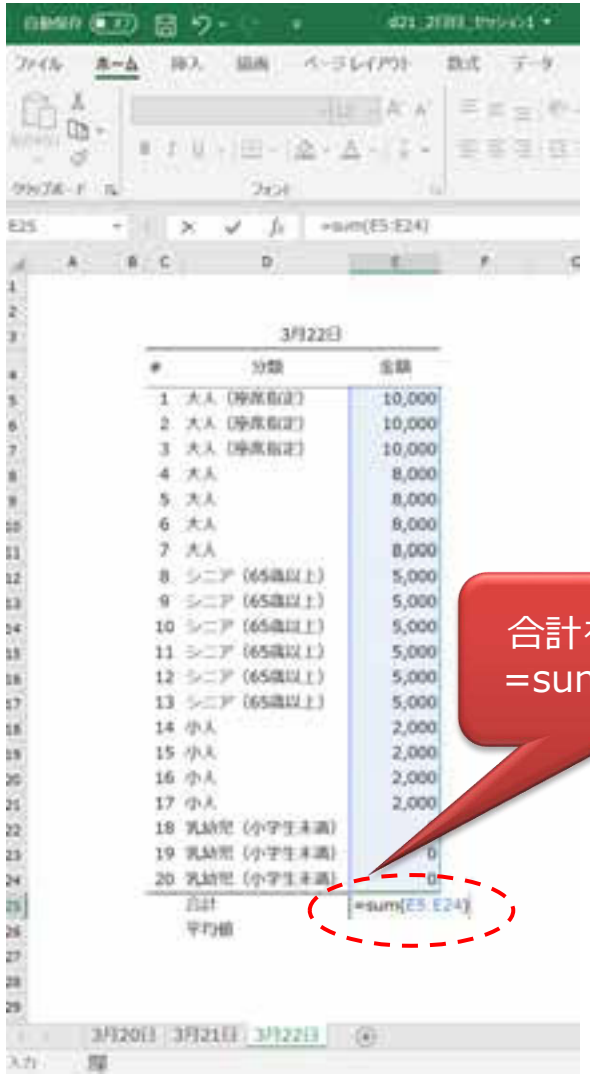


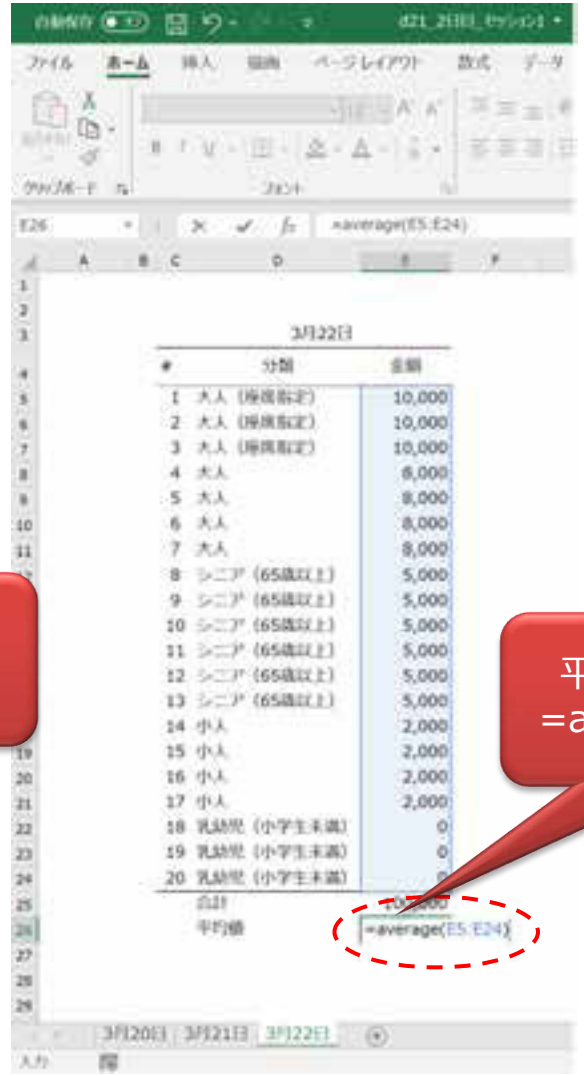
ハンズオン（1）

- ファイル名「演習データD2_1.xlsx」
- シート名「3月22日」

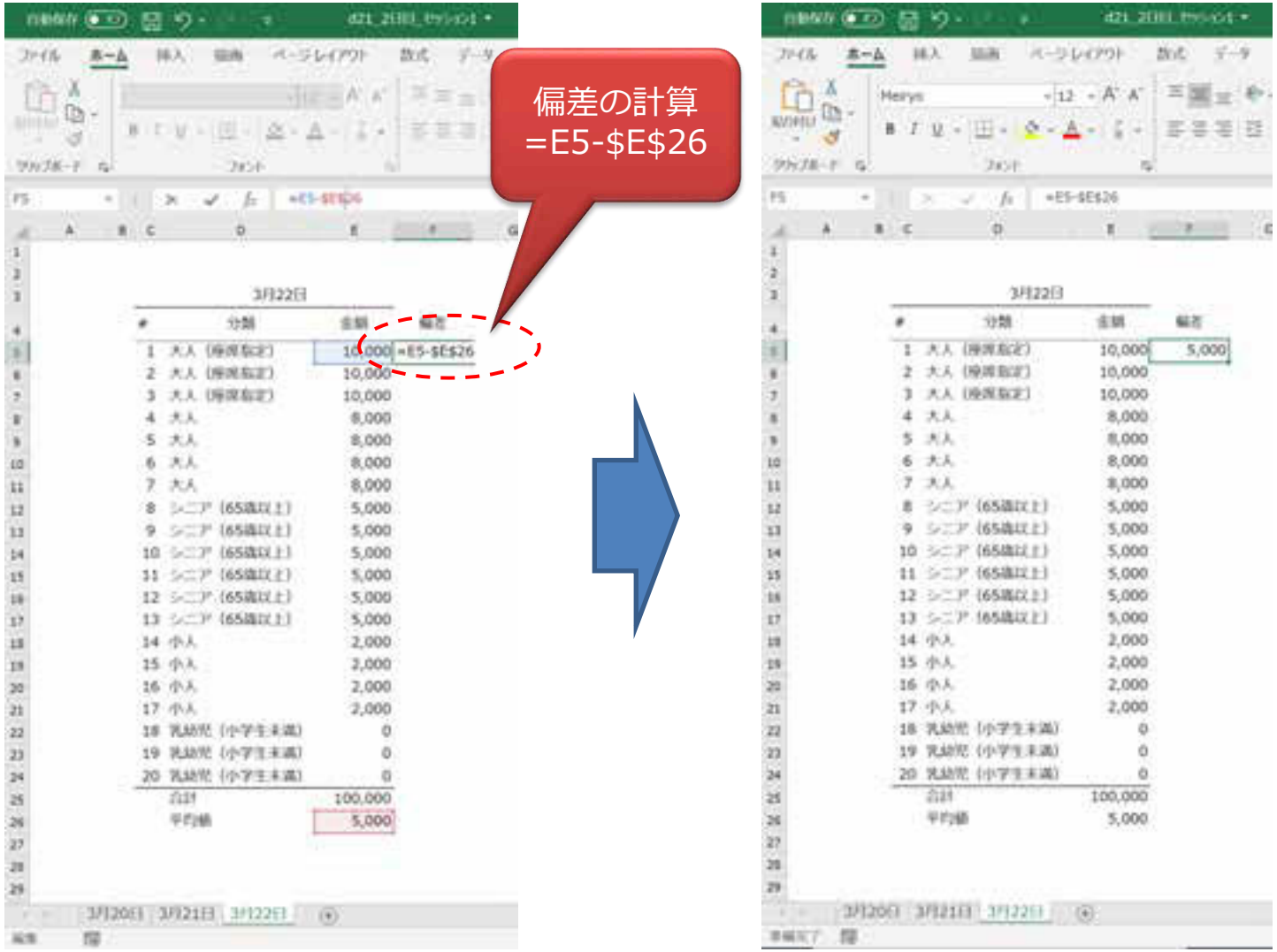


ハンズオン（2）

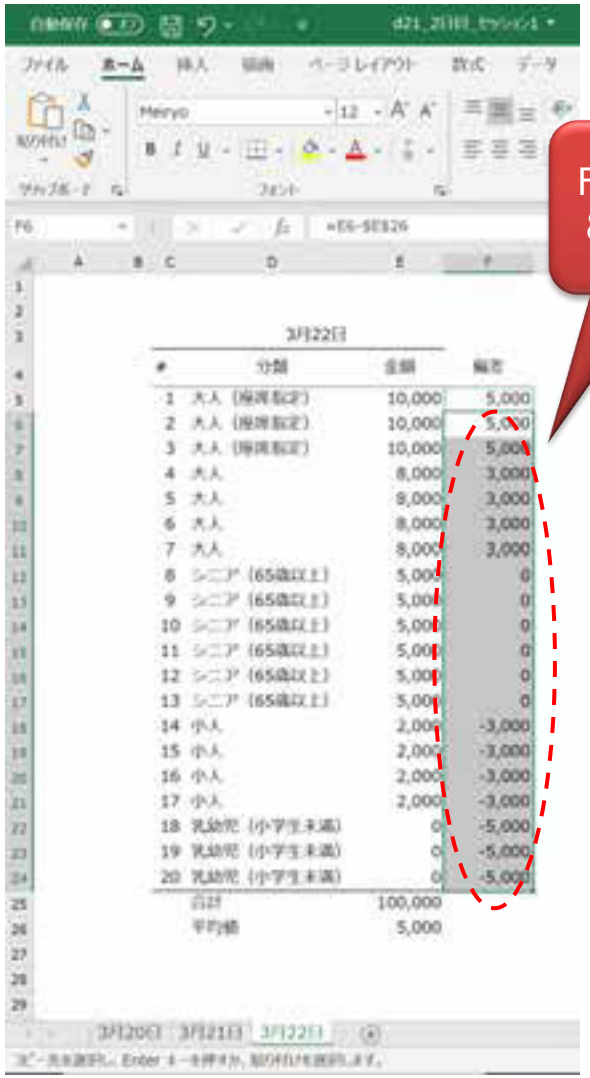


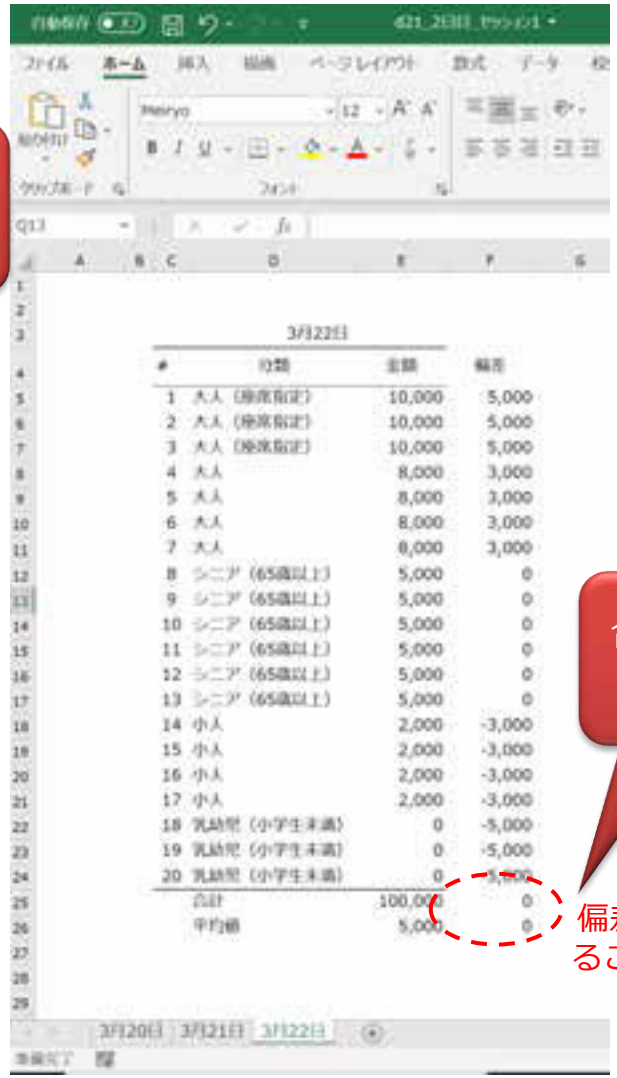


ハンズオン（3）

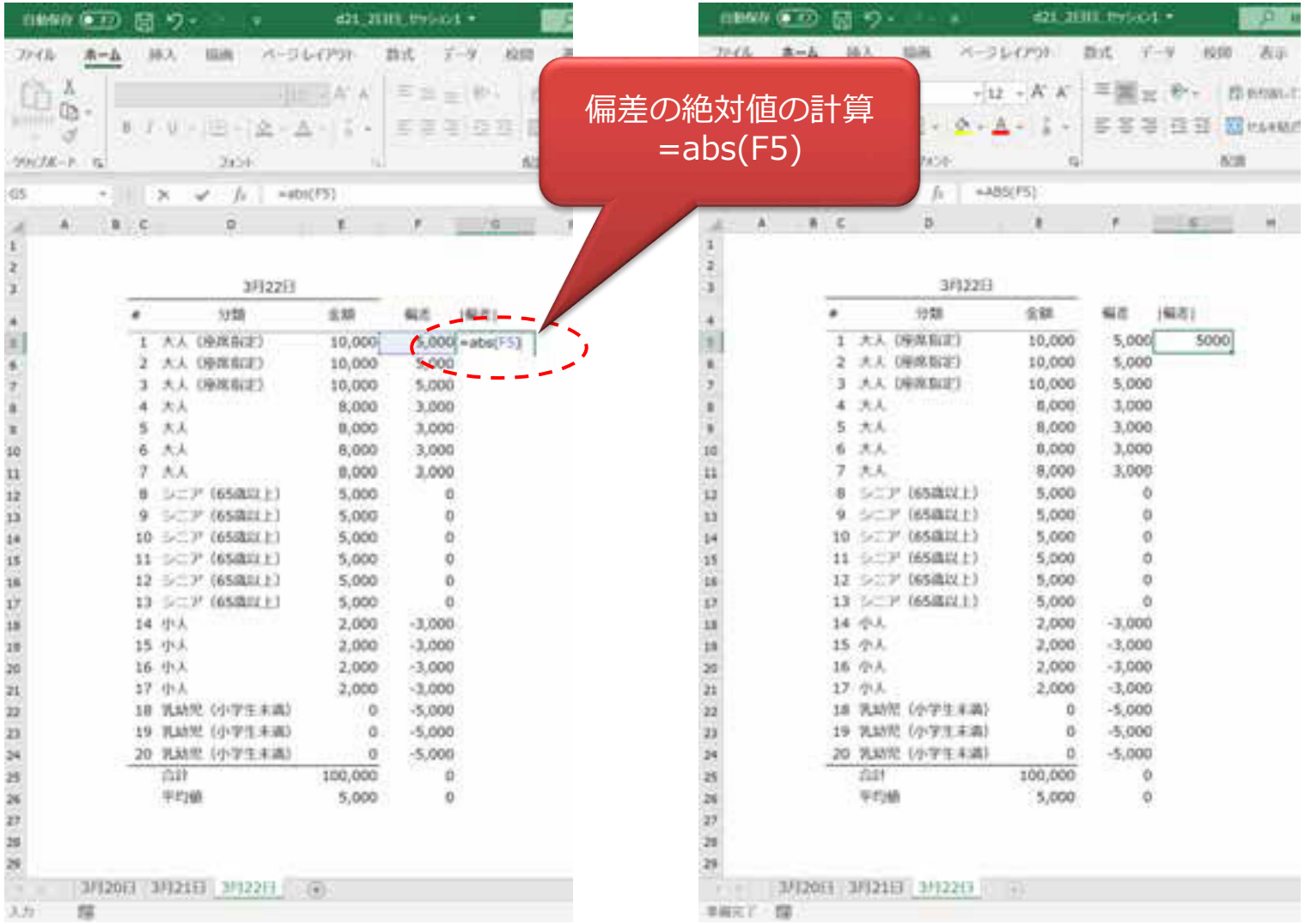


ハンズオン（4）

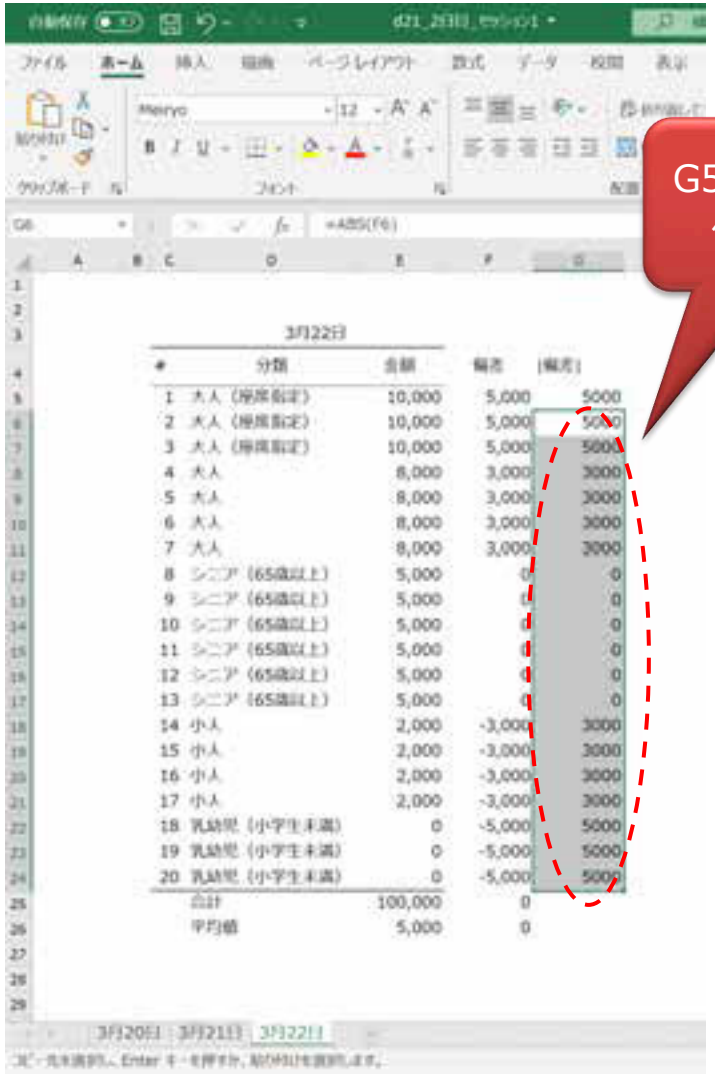

3月22日
F5をコピー
&ペースト

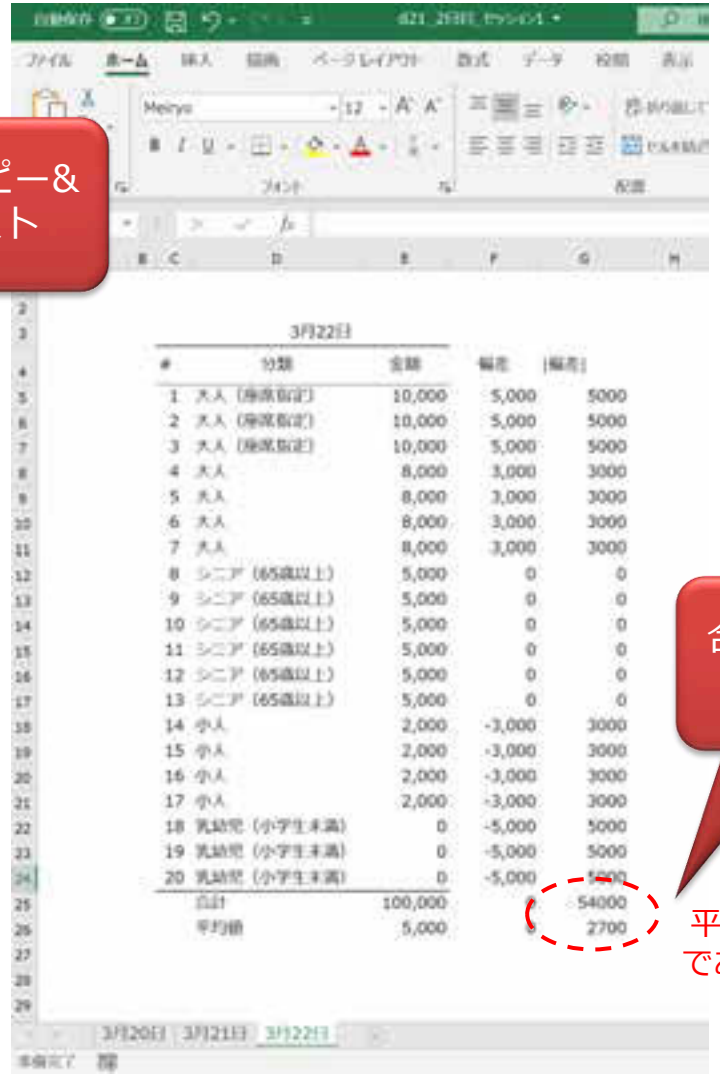

3月22日
合計と平均
値を計算
偏差の平均値がゼロにな
ることを確認しましょう。

ハンズオン（5）



ハンズオン（6）

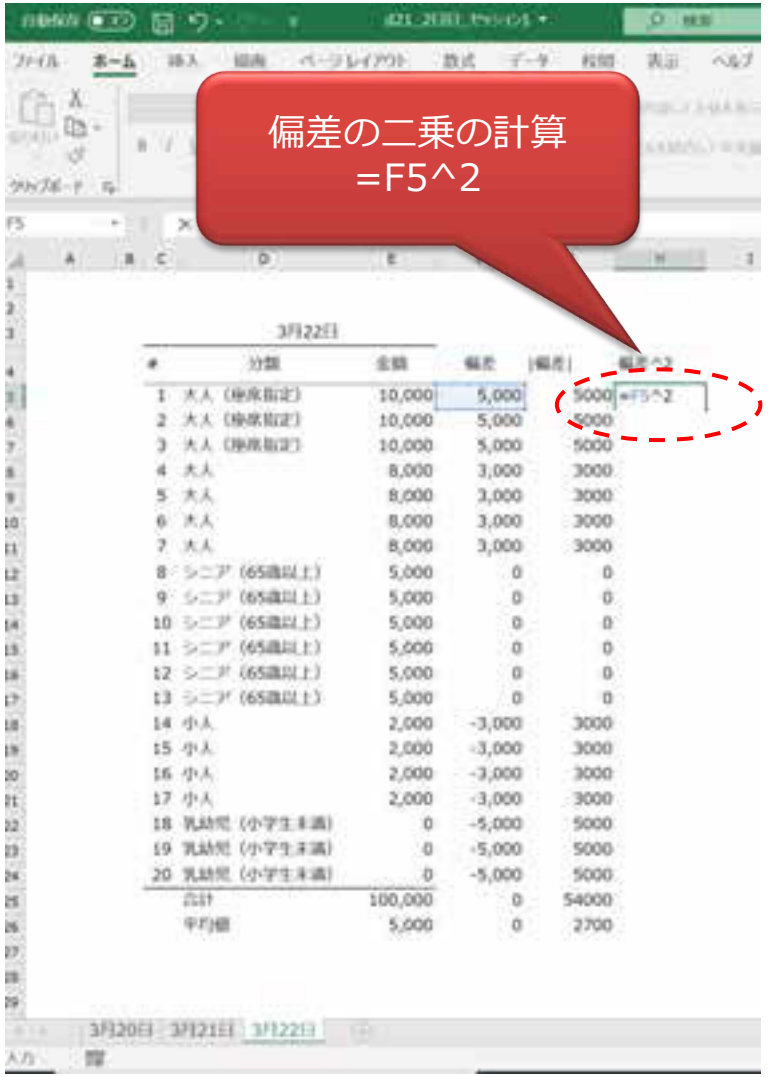


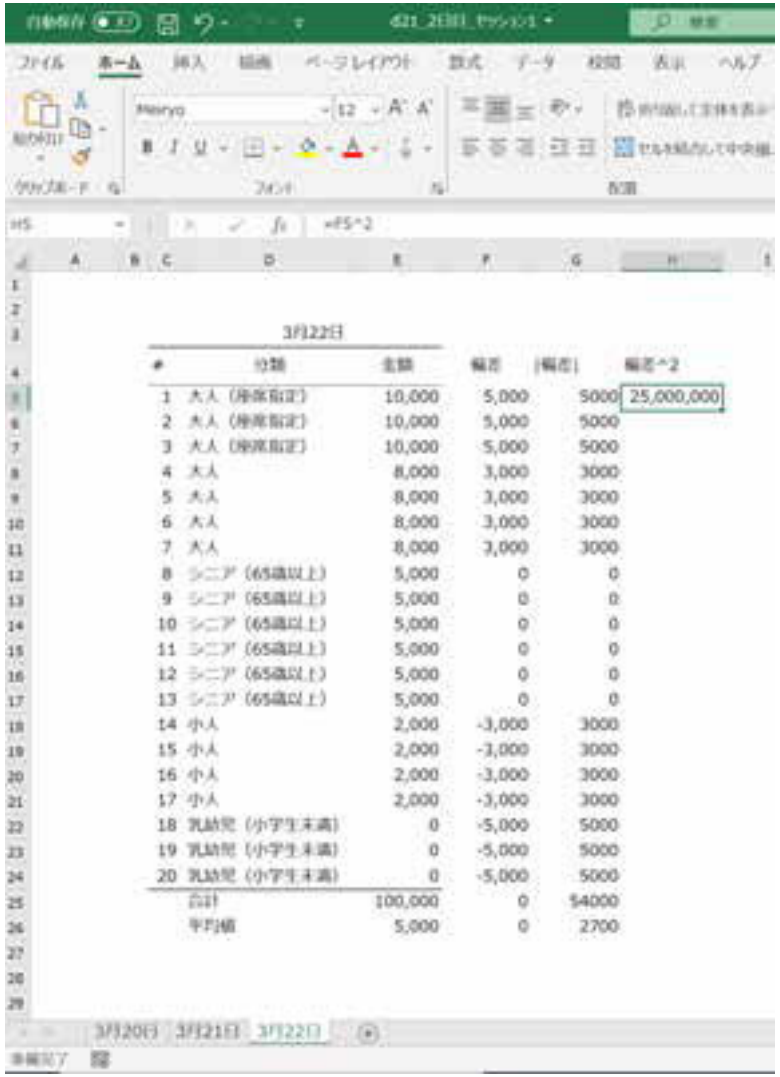


合計と平均値
を計算

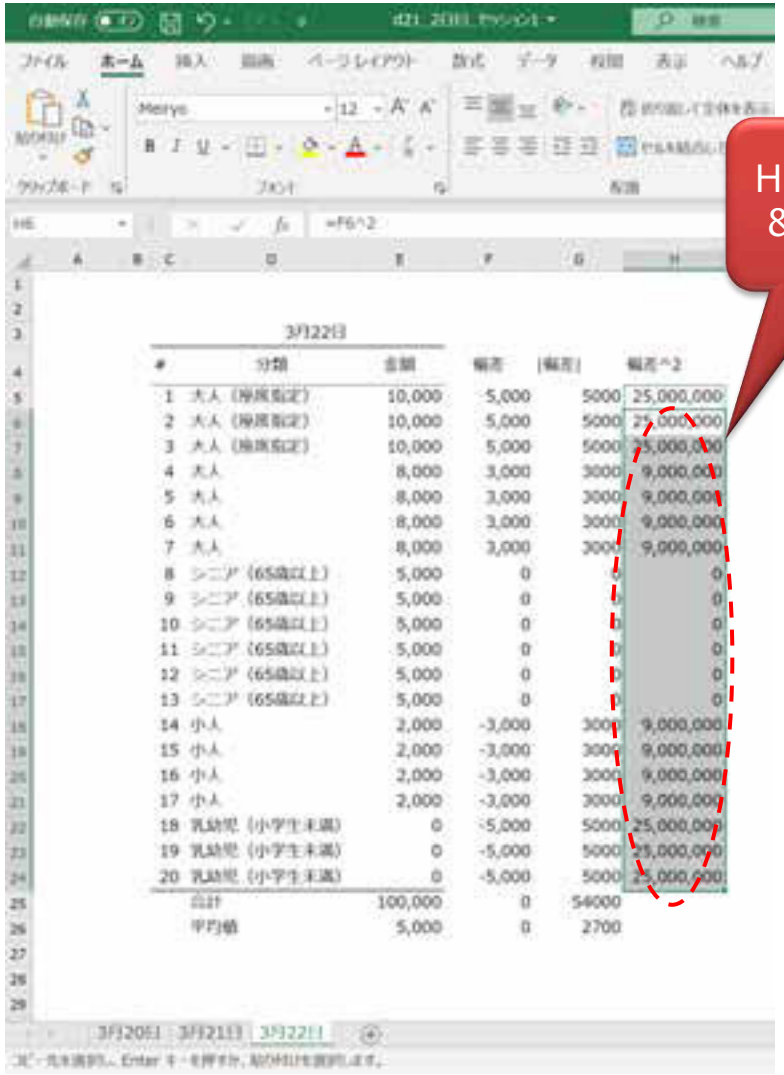
平均偏差が2700円
であることを確認し
ましょう。

ハンズオン（7）



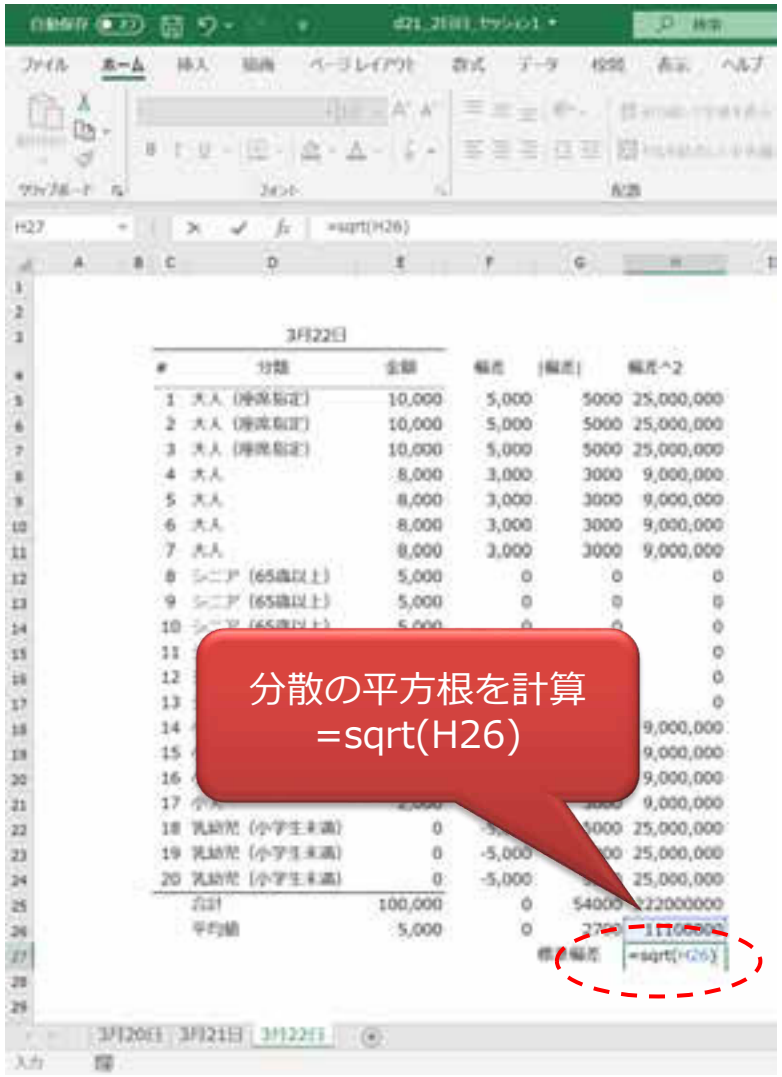


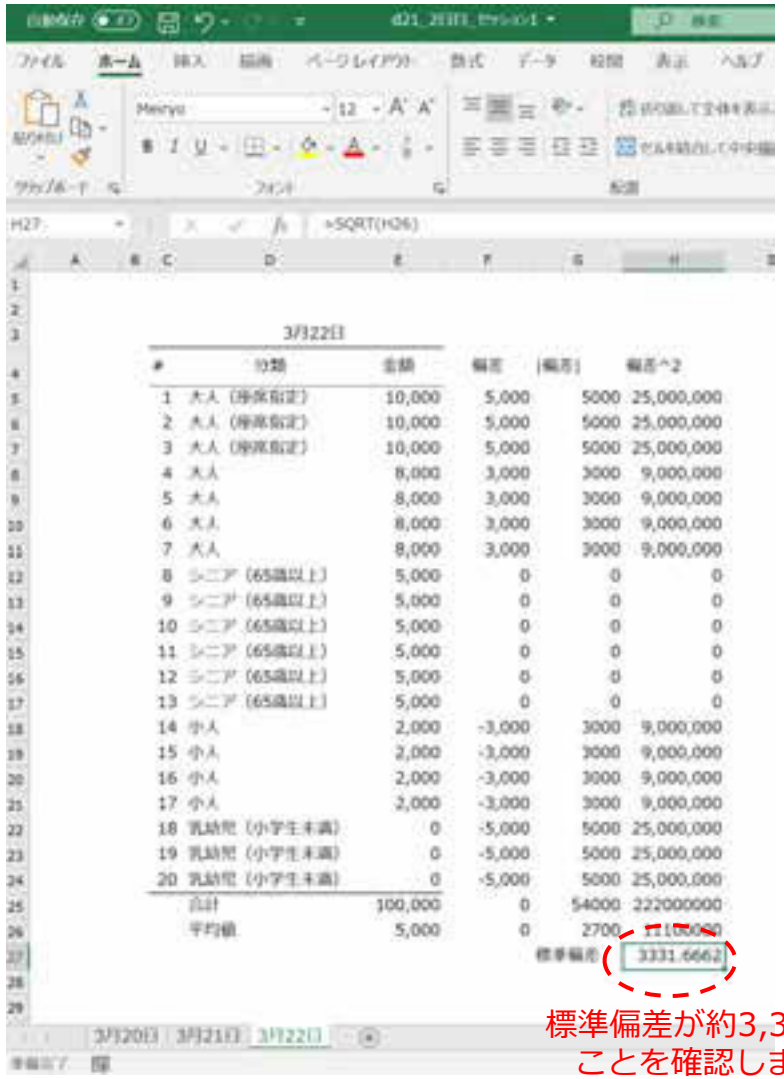
ハンズオン（8）





ハンズオン（9）

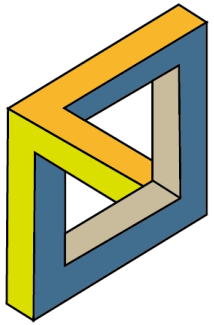




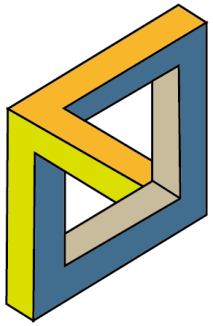
標準偏差が約3,332円である
ことを確認しましょう。

ハンズオン（10）

- シート「3月20日」、「3月21日」も同様に計算しましょう。
- 結果がどのように異なるかを確認しましょう。



データのばらつきの記述（質的変数）



データのばらつきの記述

- データのばらつきを記述するためには、「分布」という考え方が基本となります。
- 分布を調べるためには、どのようなデータがどの程度起こりやすいのか（頻度）を調べます。

度数分布表（質的データ）

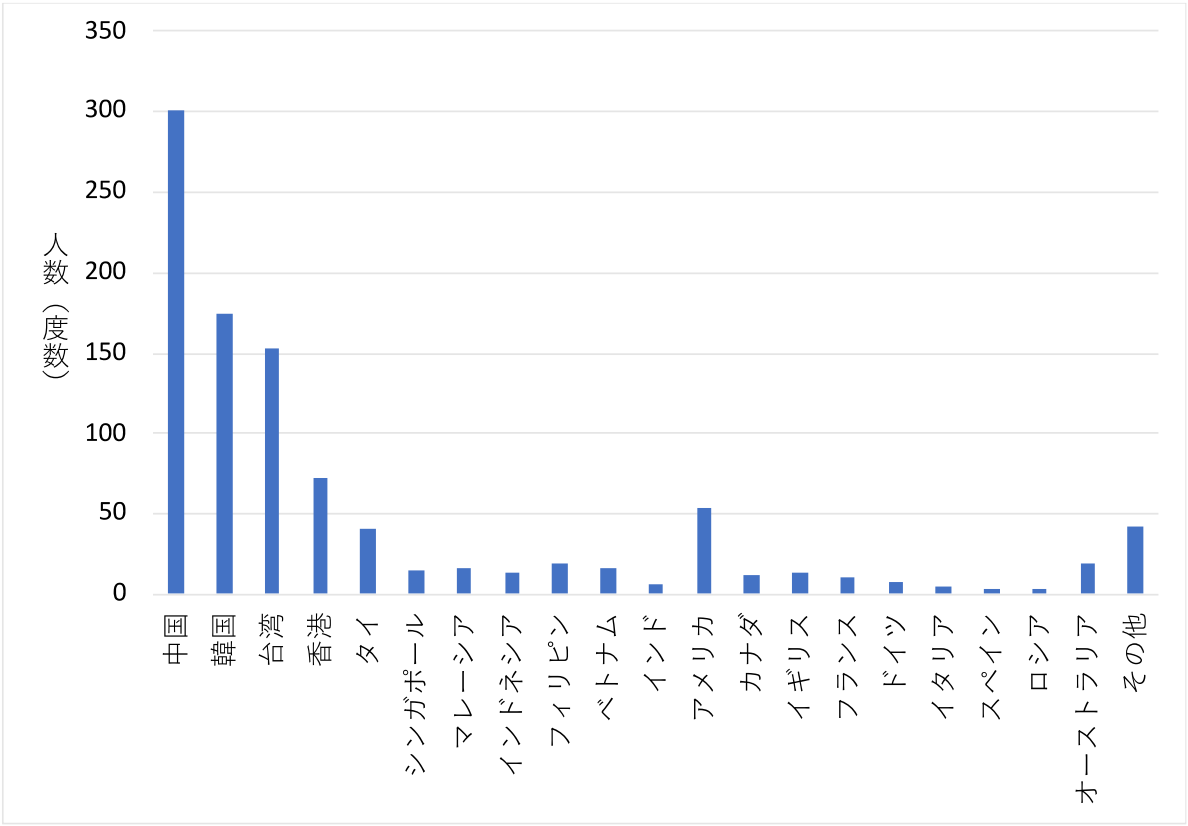
- 度数分布表はカテゴリ別にデータの頻度を集計したものです。

訪日旅行者1,000人の国別度数分布

	人数（度数）	累積度数	構成割合（相対度数）	累積構成割合（累積相対度数）
中国	301	301	30.1%	30.1%
韓国	175	476	17.5%	47.6%
台湾	153	629	15.3%	62.9%
香港	72	701	7.2%	70.1%
タイ	41	742	4.1%	74.2%
シンガポール	15	757	1.5%	75.7%
マレーシア	16	773	1.6%	77.3%
インドネシア	13	786	1.3%	78.6%
フィリピン	19	805	1.9%	80.5%
ベトナム	16	821	1.6%	82.1%
インド	6	827	0.6%	82.7%
アメリカ	54	881	5.4%	88.1%
カナダ	12	893	1.2%	89.3%
イギリス	13	906	1.3%	90.6%
フランス	11	917	1.1%	91.7%
ドイツ	8	925	0.8%	92.5%
イタリア	5	930	0.5%	93.0%
スペイン	4	934	0.4%	93.4%
ロシア	4	938	0.4%	93.8%
オーストラリア	19	957	1.9%	95.7%
その他	43	1000	4.3%	100.0%
合計	1000		100.0%	

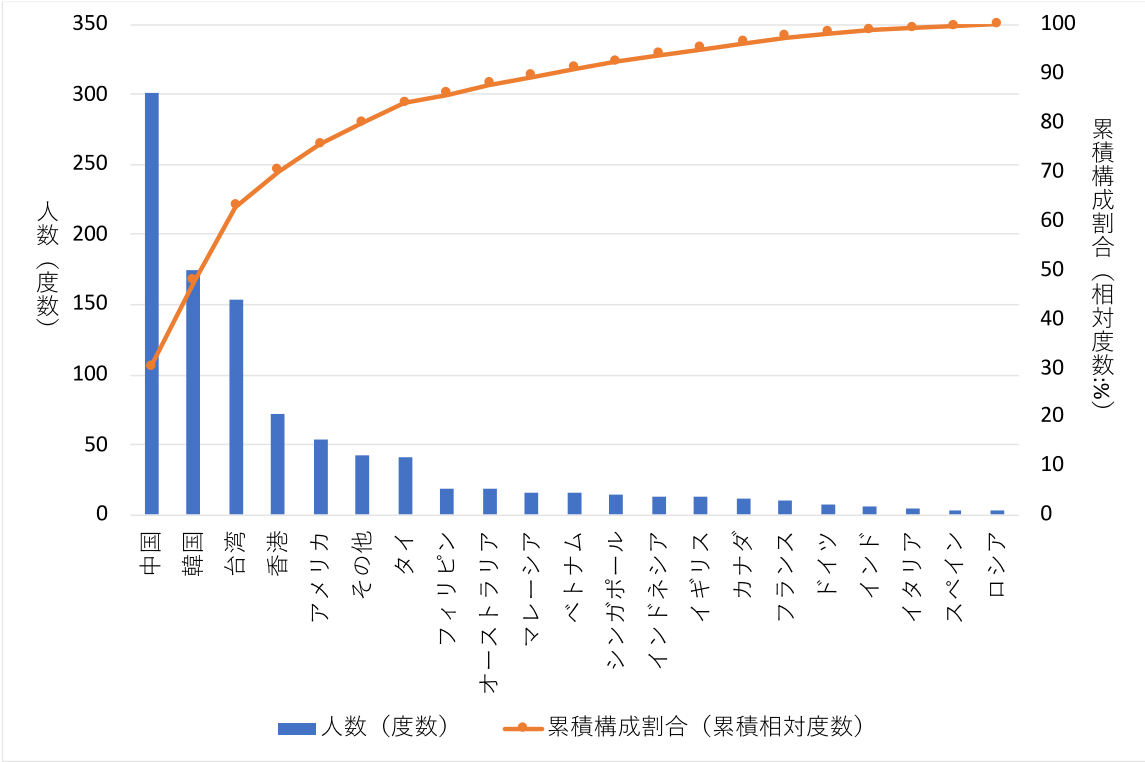
グラフ表示

- 度数分布表をグラフとして表すと以下のような形になります。



パレート図

- パレート図は、度数分布表のグラフ表示に以下の内容を追加したものになります。
  - 質的データの頻度の高い順に並べ替え
  - 累積相対度数のプロット



データのばらつきの記述（量的変数）

度数分布表（量的データ）

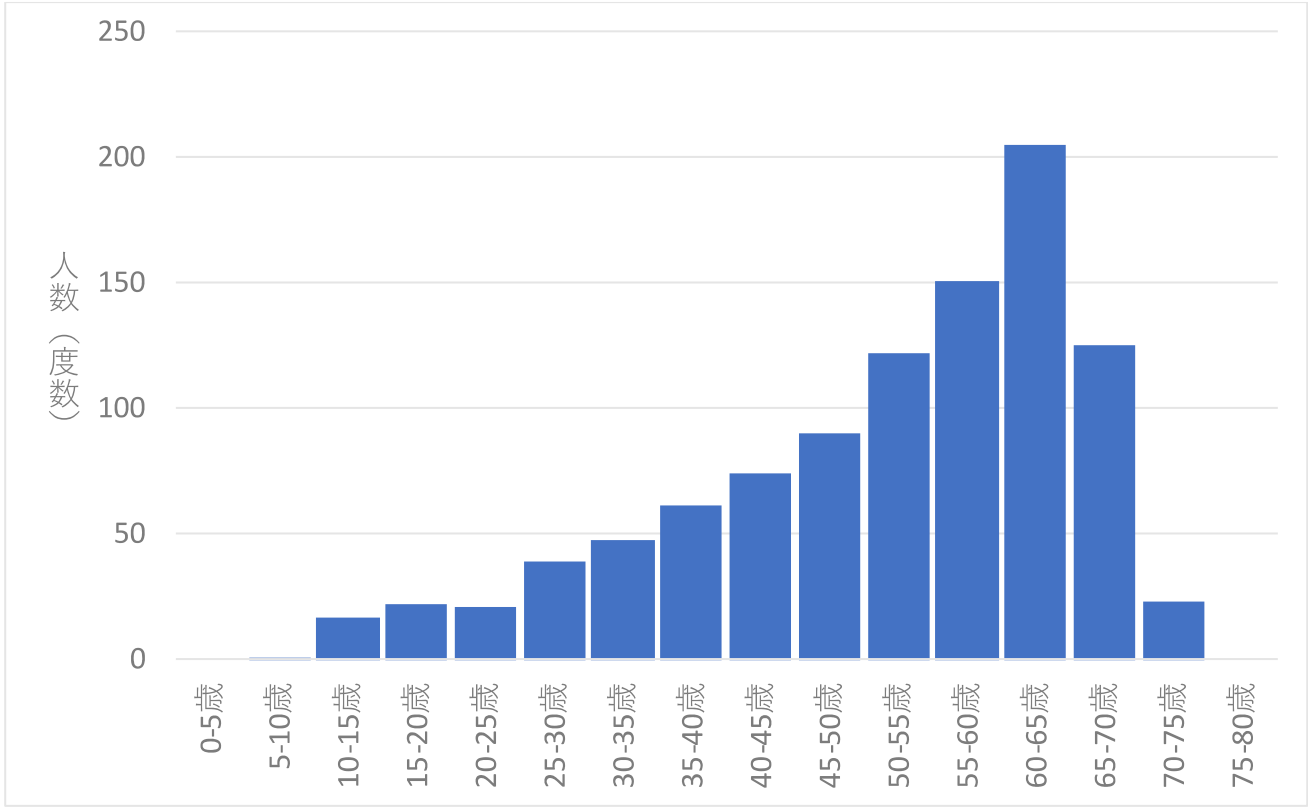
- 量的データも質的データ同様に度数分布表を作成します。
  - 質的データと異なり、階級値を設定する必要があります。

訪日旅行者1,000人の年齢分布

階級	階級値	人数（度数）	累積度数	構成割合（相対度数）	累積構成割合（累積相対度数）
0-5歳	2.5歳	0	0	0.0%	0.0%
5-10歳	7.5歳	1	1	0.1%	0.1%
10-15歳	12.5歳	17	18	1.7%	1.8%
15-20歳	17.5歳	22	40	2.2%	4.0%
20-25歳	22.5歳	21	61	2.1%	6.1%
25-30歳	27.5歳	39	100	3.9%	10.0%
30-35歳	32.5歳	48	148	4.8%	14.8%
35-40歳	37.5歳	62	210	6.2%	21.0%
40-45歳	42.5歳	74	284	7.4%	28.4%
45-50歳	47.5歳	90	374	9.0%	37.4%
50-55歳	52.5歳	122	496	12.2%	49.6%
55-60歳	57.5歳	151	647	15.1%	64.7%
60-65歳	62.5歳	205	852	20.5%	85.2%
65-70歳	67.5歳	125	977	12.5%	97.7%
70-75歳	72.5歳	23	1000	2.3%	100.0%
75-80歳	77.5歳	0	1000	0.0%	100.0%
合計		1000		100.0%	

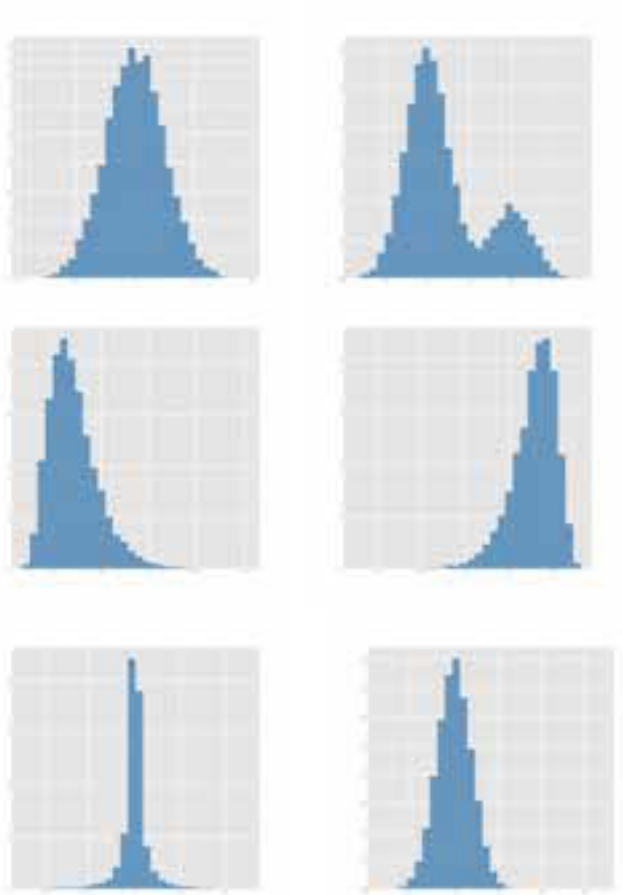
ヒストグラム

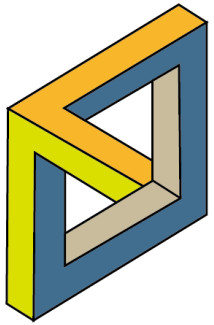
- 量的データの度数分布表をグラフにしたものを「ヒストグラム」と呼びます。



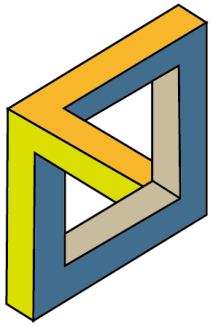
ヒストグラムのチェックポイント

- 単峰型か複峰型か？
- 左右の歪みはあるか？
- 上下の尖り・外れ値はあるか？



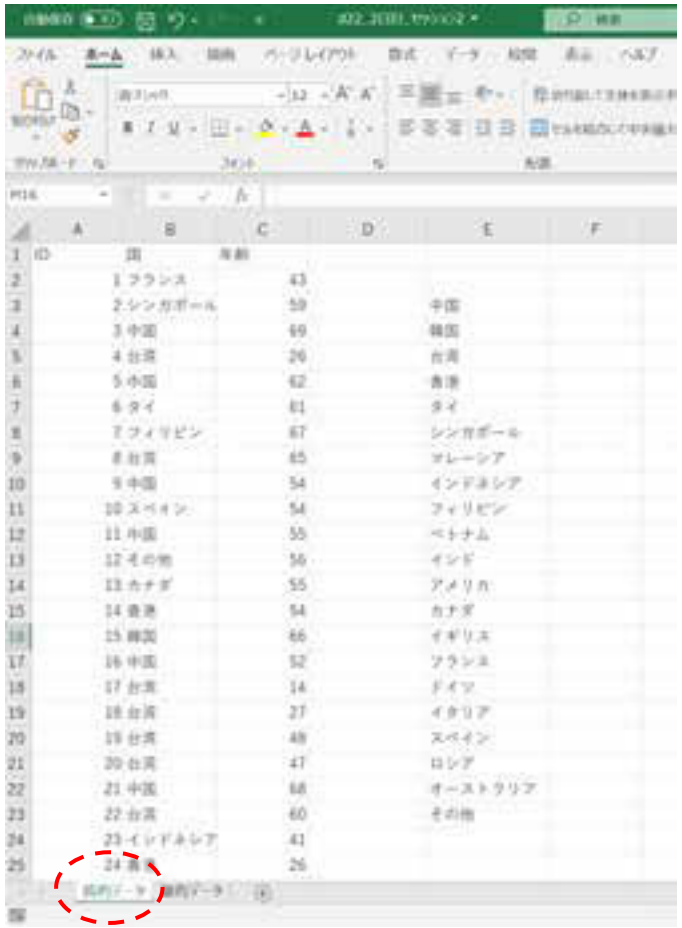


ハンズオン



ハンズオン（1）

- ファイル名「演習データD2_2.xlsx」
- ワークシート「質的データ」



ハンズオン（2）

F3 =countif($B$2:$B$1001,E3)

	A	B	C	D	E	F	G	H	I
1	ID	国	年齢						
2	1	フランス	43			人数（度数）			
3	2	シンガポール	59		中国	=countif($B$2:$B$1001,E3)			
4	3	中国	69		韓国				
5	4	台湾	26		台湾				
6	5	中国	62		香港				
7	6	タイ	61		タイ				
8	7	フィリピン	67		シンガポール				
9	8	台湾	65		マレーシア				
10	9	中国	54		インドネシア				
11	10	スペイン	54		フィリピン				
12	11	中国	55		ベトナム				
13	12	その他	56		インド				
14	13	カナダ	55		アメリカ				
15	14	香港	54		カナダ				
16	15	韓国	66		イギリス				
17	16	中国	52		フランス				
18	17	台湾	14		ドイツ				
19	18	台湾	27		イタリア				
20	19	台湾	48		スペイン				
21	20	台湾	47		ロシア				
22	21	中国	68		オーストラリア				
23	22	台湾	60		その他				
24	23	インドネシア	41						

国名と一致するデータ数を数える
=countif($B$2:$B$1001,E3)

ハンズオン（3）

F3 =COUNTIF($B$2:$B$1001,E3)

	A	B	C	D	E	F	G	H	I
1	ID	国	年齢						
2	1	フランス	43			人数（度数）			
3	2	シンガポール	59		中国	301			
4	3	中国	69		韓国				
5	4	台湾	26		台湾				
6	5	中国	62		香港				
7	6	タイ	61		タイ				
8	7	フィリピン	67		シンガポール				
9	8	台湾	65		マレーシア				
10	9	中国	54		インドネシア				
11	10	スペイン	54		フィリピン				
12	11	中国	55		ベトナム				
13	12	その他	56		インド				
14	13	カナダ	55		アメリカ				
15	14	香港	54		カナダ				
16	15	韓国	66		イギリス				
17	16	中国	52		フランス				
18	17	台湾	14		ドイツ				
19	18	台湾	27		イタリア				
20	19	台湾	48		スペイン				
21	20	台湾	47		ロシア				
22	21	中国	68		オーストラリア				
23	22	台湾	60		その他				
24	23	インドネシア	41						

中国本土から訪日した人の
人数は301人

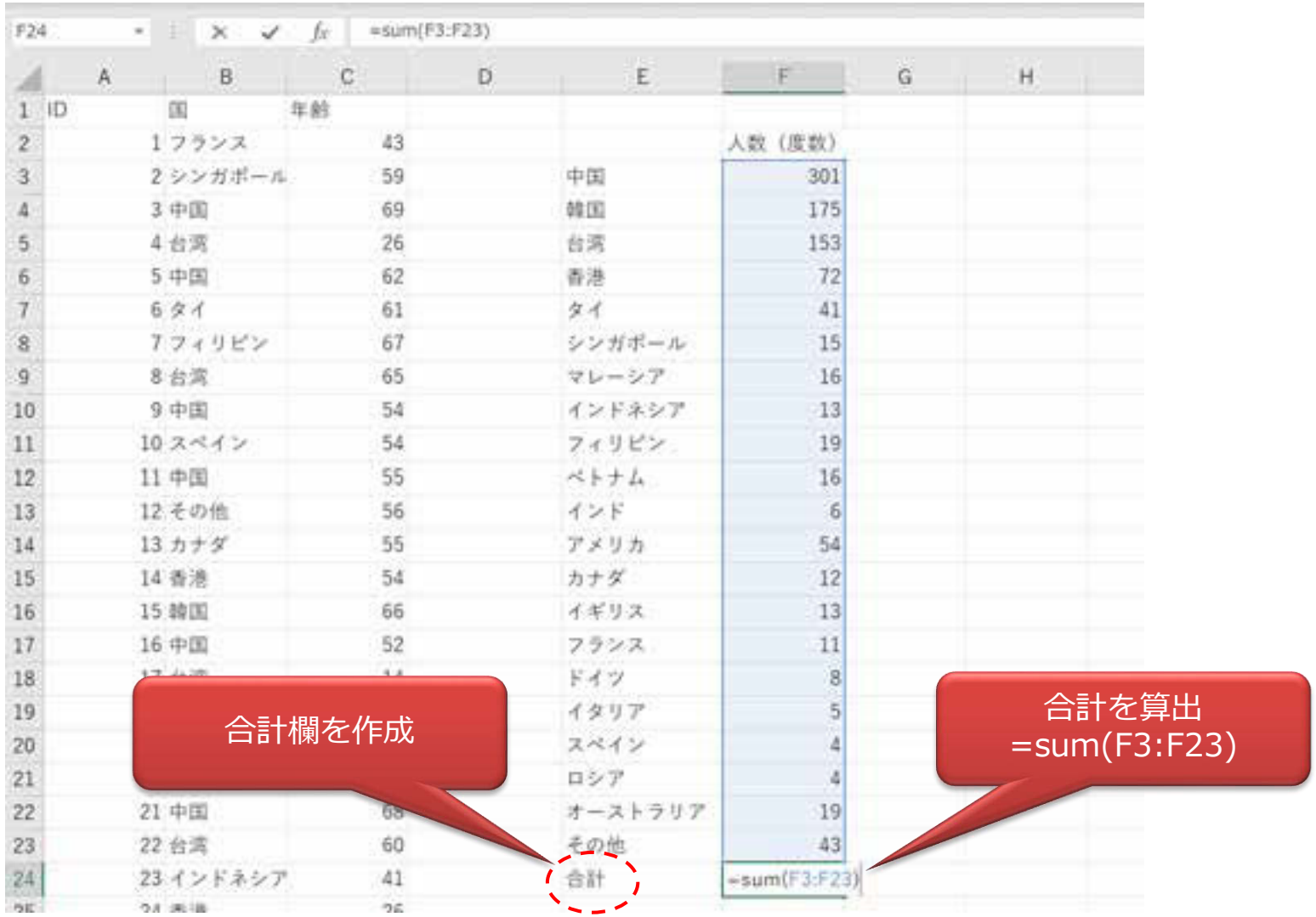
ハンズオン（4）

F4 =COUNTIF($B$2:$B$1001,E4)

	A	B	C	D	E	F	G	H
1	ID	国	年齢					
2	1	フランス	43			人数（度数）		
3	2	シンガポール	59		中国	301		
4	3	中国	69		韓国	175		
5	4	台湾	26		台湾	153		
6	5	中国	62		香港	72		
7	6	タイ	61		タイ	41		
8	7	フィリピン	67		シンガポール	15		
9	8	台湾	65		マレーシア	16		
10	9	中国	54		インドネシア	13		
11	10	スペイン	54		フィリピン	19		
12	11	中国	55		ベトナム	16		
13	12	その他	56		インド	6		
14	13	カナダ	55		アメリカ	54		
15	14	香港	54		カナダ	12		
16	15	韓国	66		イギリス	13		
17	16	中国	52		フランス	11		
18	17	台湾	14		ドイツ	8		
19	18	台湾	27		イタリア	5		
20	19	台湾	48		スペイン	4		
21	20	台湾	47		ロシア	4		
22	21	中国	68		オーストラリア	19		
23	22	台湾	60		その他	43		
24	23	インドネシア	41					

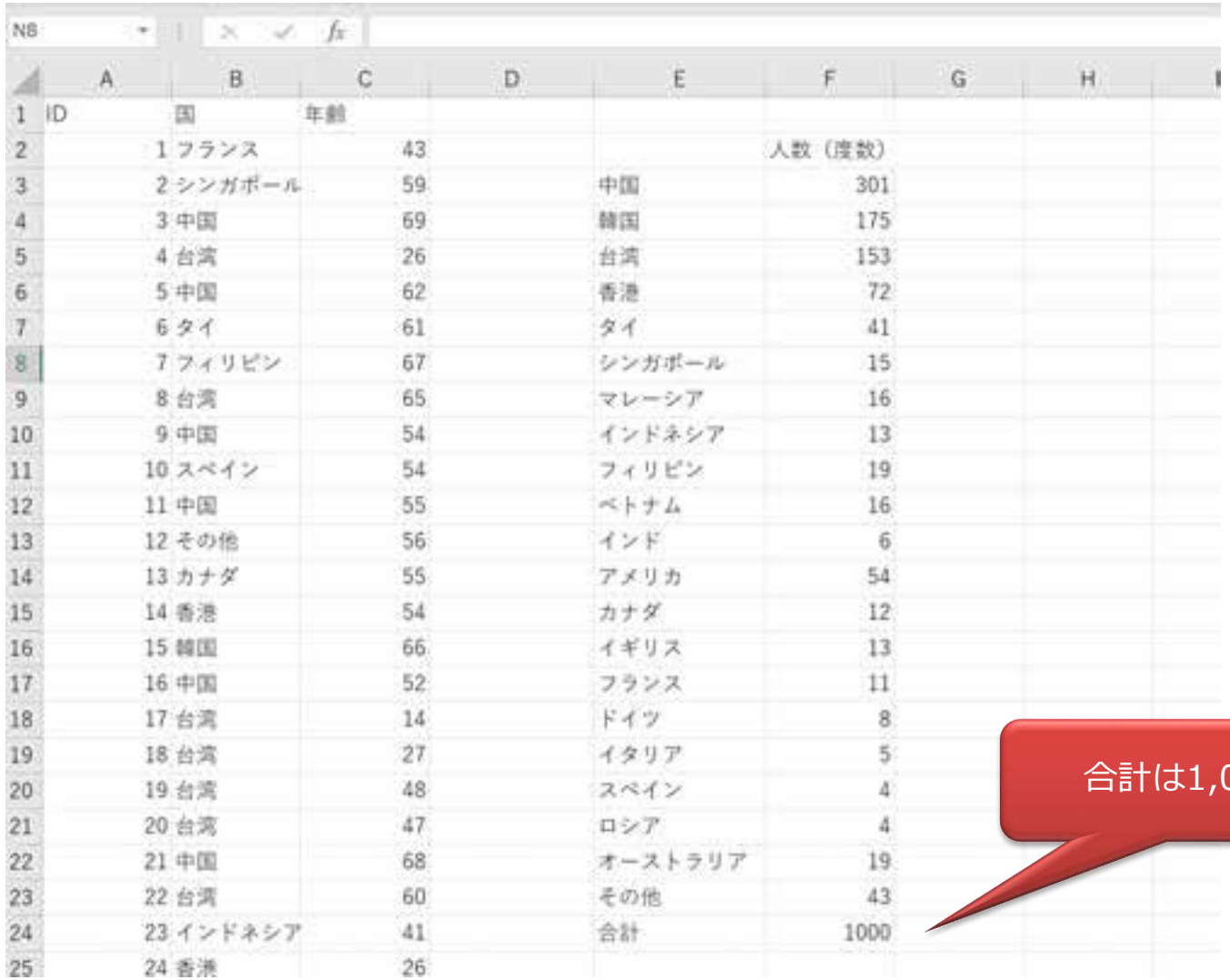
コピー&ペースト

ハンズオン（5）



	A	B	C	D	E	F	G	H
1	ID	国	年齢					
2		1 フランス	43			人数（度数）		
3		2 シンガポール	59		中国	301		
4		3 中国	69		韓国	175		
5		4 台湾	26		台湾	153		
6		5 中国	62		香港	72		
7		6 タイ	61		タイ	41		
8		7 フィリピン	67		シンガポール	15		
9		8 台湾	65		マレーシア	16		
10		9 中国	54		インドネシア	13		
11		10 スペイン	54		フィリピン	19		
12		11 中国	55		ベトナム	16		
13		12 その他	56		インド	6		
14		13 カナダ	55		アメリカ	54		
15		14 香港	54		カナダ	12		
16		15 韓国	66		イギリス	13		
17		16 中国	52		フランス	11		
18					ドイツ	8		
19					イタリア	5		
20					スペイン	4		
21					ロシア	4		
22		21 中国	68		オーストラリア	19		
23		22 台湾	60		その他	43		
24		23 インドネシア	41		合計	=sum(F3:F23)		

ハンズオン（6）



ID	国	年齢			
1	フランス	43		人数（度数）	
2	シンガポール	59	中国	301	
3	中国	69	韓国	175	
4	台湾	26	台湾	153	
5	中国	62	香港	72	
6	タイ	61	タイ	41	
7	フィリピン	67	シンガポール	15	
8	台湾	65	マレーシア	16	
9	中国	54	インドネシア	13	
10	スペイン	54	フィリピン	19	
11	中国	55	ベトナム	16	
12	その他	56	インド	6	
13	カナダ	55	アメリカ	54	
14	香港	54	カナダ	12	
15	韓国	66	イギリス	13	
16	中国	52	フランス	11	
17	台湾	14	ドイツ	8	
18	台湾	27	イタリア	5	
19	台湾	48	スペイン	4	
20	台湾	47	ロシア	4	
21	中国	68	オーストラリア	19	
22	台湾	60	その他	43	
23	インドネシア	41	合計	1000	
24	香港	26			

合計は1,000人

ハンズオン（7）



F3 =F3

	A	B	C	D	E	F	G
1	ID	国	年齢				
2		1 フランス	43			人数（度数）	累積度数
3		2 シンガポール	59		中国	301	=F3
4		3 中国	69		韓国	175	
5		4 台湾	26		台湾	153	
6		5 中国	62		香港	72	
7		6 タイ	61		タイ	41	
8		7 フィリピン	67		シンガポール	15	
9		8 台湾	65		マレーシア	16	
10		9 中国	54		インドネシア	13	
11		10 スペイン	54		フィリピン	19	
12		11 中国	55		ベトナム	16	
13		12 その他	56		インド	6	
14		13 カナダ	55		アメリカ	54	
15		14 香港	54		カナダ	12	
16		15 韓国	66		イギリス	13	
17		16 中国	52		フランス	11	
18		17 台湾	14		ドイツ	8	
19		18 台湾	27		イタリア	5	
20		19 台湾	48		スペイン	4	
21		20 台湾	47		ロシア	4	
22		21 中国	68		オーストラリア	19	
23		22 台湾	60		その他	43	
24		23 インドネシア	41		合計	1000	

累積度数の欄を作成

1行目は度数の値を入れる。
=F3

ハンズオン（8）

F4 =G3+F4

	A	B	C	D	E	F	G	H
1	ID	国	年齢					
2	1	フランス	43			人数（度数）	累積度数	
3	2	シンガポール	59		中国	301	301	
4	3	中国	69		韓国	175	=G3+F4	
5	4	台湾	26		台湾	153		
6	5	中国	62		香港	72		
7	6	タイ	61		タイ	41		
8	7	フィリピン	67		シンガポール	15		
9	8	台湾	65		マレーシア	16		
10	9	中国	54		インドネシア	13		
11	10	スペイン	54		フィリピン	19		
12	11	中国	55		ベトナム	16		
13	12	その他	56		インド	6		
14	13	カナダ	55		アメリカ	54		
15	14	香港	54		カナダ	12		
16	15	韓国	66		イギリス	13		
17	16	中国	52		フランス	11		
18	17	台湾	14		ドイツ	8		
19	18	台湾	27		イタリア	5		
20	19	台湾	48		スペイン	4		
21	20	台湾	47		ロシア	4		
22	21	中国	68		オーストラリア	19		
23	22	台湾	60		その他	43		
24	23	インドネシア	41		合計	1000		
25	24	香港	26					

2行目以降は前の行の合計と度数を足し合わせる。

ハンズオン（9）

G5 =G4+F5

	A	B	C	D	E	F	G	H
1	ID	国	年齢					
2		1 フランス	43			人数（度数）	累積度数	
3		2 シンガポール	59		中国	301	301	
4		3 中国	69		韓国	175	476	
5		4 台湾	26		台湾	153	629	
6		5 中国	62		香港	72	701	
7		6 タイ	61		タイ	41	742	
8		7 フィリピン	67		シンガポール	15	757	
9		8 台湾	65		マレーシア	16	773	
10		9 中国	54		インドネシア	13	786	
11		10 スペイン	54		フィリピン	19	805	
12		11 中国	55		ベトナム	16	821	
13		12 その他	56		インド	6	827	
14		13 カナダ	55		アメリカ	54	881	
15		14 香港	54		カナダ	12	893	
16		15 韓国	66		イギリス	13	906	
17		16 中国	52		フランス	11	917	
18		17 台湾	14		ドイツ	8	925	
19		18 台湾	27		イタリア	5	930	
20		19 台湾	48		スペイン	4	934	
21		20 台湾	47		ロシア	4	938	
22		21 中国	68		オーストラリア	19	957	
23		22 台湾	60		その他	43	1000	
24		23 インドネシア	41		合計	1000		

コピー&ペースト

ハンズオン（10）



	A	B	C	D	E	F	G	H
1	ID	国	年齢			人数（度数）	累積度数	相対度数
2	1	フランス	43					
3	2	シンガポール	59		中国	301	301	=F3/$F$24
4	3	中国	69		韓国	175	476	
5	4	台湾	26		台湾	153	629	
6	5	中国	62		香港	72	701	
7	6	タイ	61		タイ	41	742	
8	7	フィリピン	67		シンガポール	15	757	
9	8	台湾	65		マレーシア	16	773	
10	9	中国	54		インドネシア	13	786	
11	10	スペイン	54		フィリピン	19	805	
12	11	中国	55		ベトナム	16	821	
13	12	その他	56		インド	6	827	
14	13	カナダ	55		アメリカ	54	881	
15	14	香港	54		カナダ	12	893	
16	15	韓国	66		イギリス	13	906	
17	16	中国	52		フランス	11	917	
18	17	台湾	14		ドイツ	8	925	
19	18	台湾	27		イタリア	5	930	
20	19	台湾	48		スペイン	4	934	
21	20	台湾	47		ロシア	4	938	
22	21	中国	68		オーストラリア	19	957	
23	22	台湾	60		その他	43	1000	
24	23	インドネシア	41		合計	1000		
25	24	香港	26					

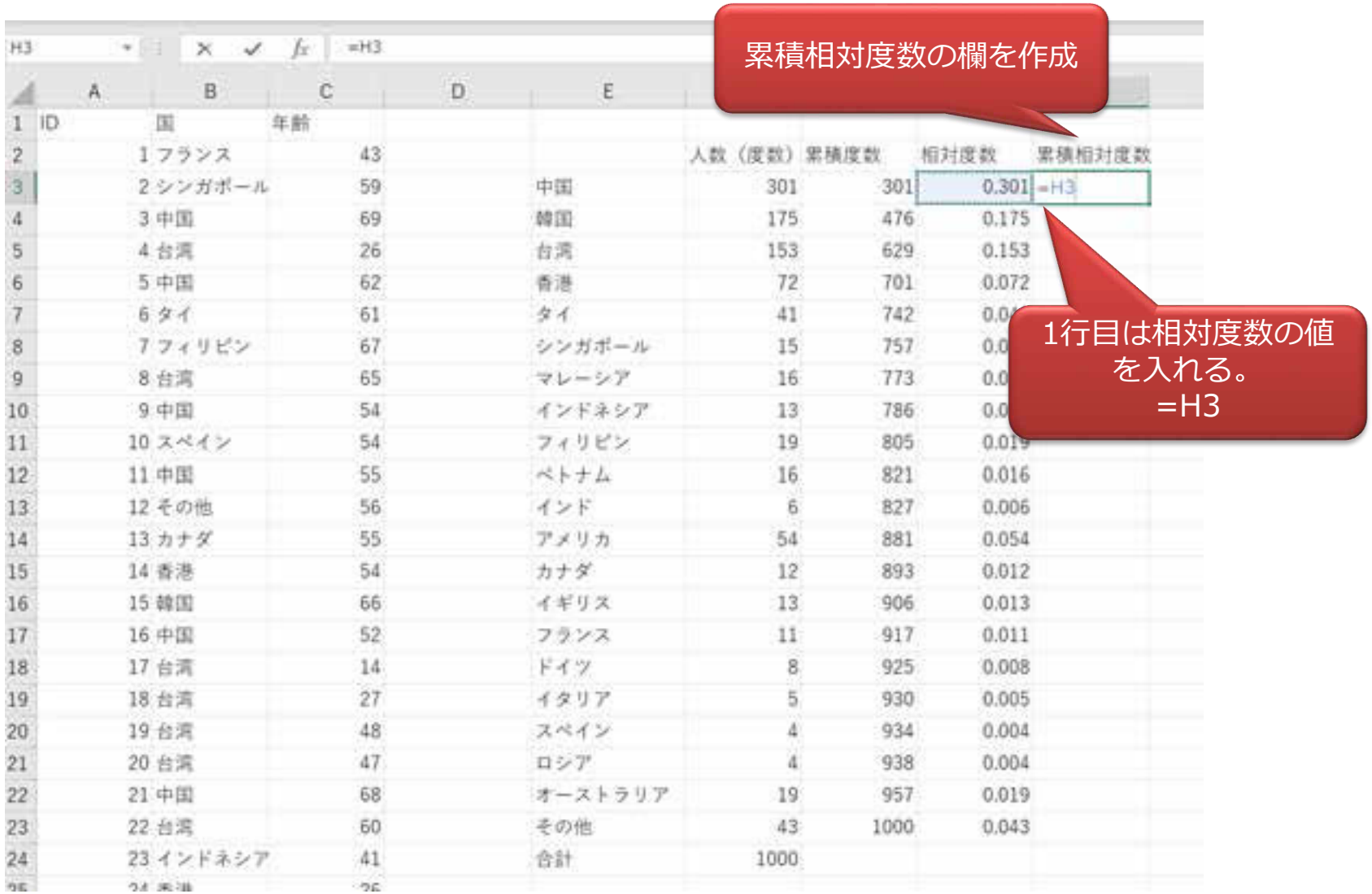
ハンズオン（11）

H4 =F4/$F$24

	A	B	C	D	E	F	G	H
1	ID	国	年齢					
2		1 フランス	43			人数（度数）	累積度数	相対度数
3		2 シンガポール	59		中国	301	301	0.301
4		3 中国	69		韓国	175	476	0.175
5		4 台湾	26		台湾	153	629	0.153
6		5 中国	62		香港	72	701	0.07
7		6 タイ	61		タイ	41	742	0.041
8		7 フィリピン	67		シンガポール	15	757	0.
9		8 台湾	65		マレーシア	16	773	0.
10		9 中国	54		インドネシア	13	786	0.013
11		10 スペイン	54		フィリピン	19	805	0.019
12		11 中国	55		ベトナム	16	821	0.016
13		12 その他	56		インド	6	827	0.006
14		13 カナダ	55		アメリカ	54	881	0.054
15		14 香港	54		カナダ	12	893	0.012
16		15 韓国	66		イギリス	13	906	0.013
17		16 中国	52		フランス	11	917	0.011
18		17 台湾	14		ドイツ	8	925	0.008
19		18 台湾	27		イタリア	5	930	0.005
20		19 台湾	48		スペイン	4	934	0.004
21		20 台湾	47		ロシア	4	938	0.004
22		21 中国	68		オーストラリア	19	957	0.019
23		22 台湾	60		その他	43	1000	0.043
24		23 インドネシア	41		合計	1000		
25		24 香港	26					

コピー&ペースト

ハンズオン（12）



H3 =H3

	A	B	C	D	E	F	G	H	I
1	ID	国	年齢						
2		1 フランス	43			人数（度数）	累積度数	相対度数	累積相対度数
3		2 シンガポール	59		中国	301	301	0.301	=H3
4		3 中国	69		韓国	175	476	0.175	
5		4 台湾	26		台湾	153	629	0.153	
6		5 中国	62		香港	72	701	0.072	
7		6 タイ	61		タイ	41	742	0.0	
8		7 フィリピン	67		シンガポール	15	757	0.0	
9		8 台湾	65		マレーシア	16	773	0.0	
10		9 中国	54		インドネシア	13	786	0.0	
11		10 スペイン	54		フィリピン	19	805	0.019	
12		11 中国	55		ベトナム	16	821	0.016	
13		12 その他	56		インド	6	827	0.006	
14		13 カナダ	55		アメリカ	54	881	0.054	
15		14 香港	54		カナダ	12	893	0.012	
16		15 韓国	66		イギリス	13	906	0.013	
17		16 中国	52		フランス	11	917	0.011	
18		17 台湾	14		ドイツ	8	925	0.008	
19		18 台湾	27		イタリア	5	930	0.005	
20		19 台湾	48		スペイン	4	934	0.004	
21		20 台湾	47		ロシア	4	938	0.004	
22		21 中国	68		オーストラリア	19	957	0.019	
23		22 台湾	60		その他	43	1000	0.043	
24		23 インドネシア	41		合計	1000			

ハンズオン（13）

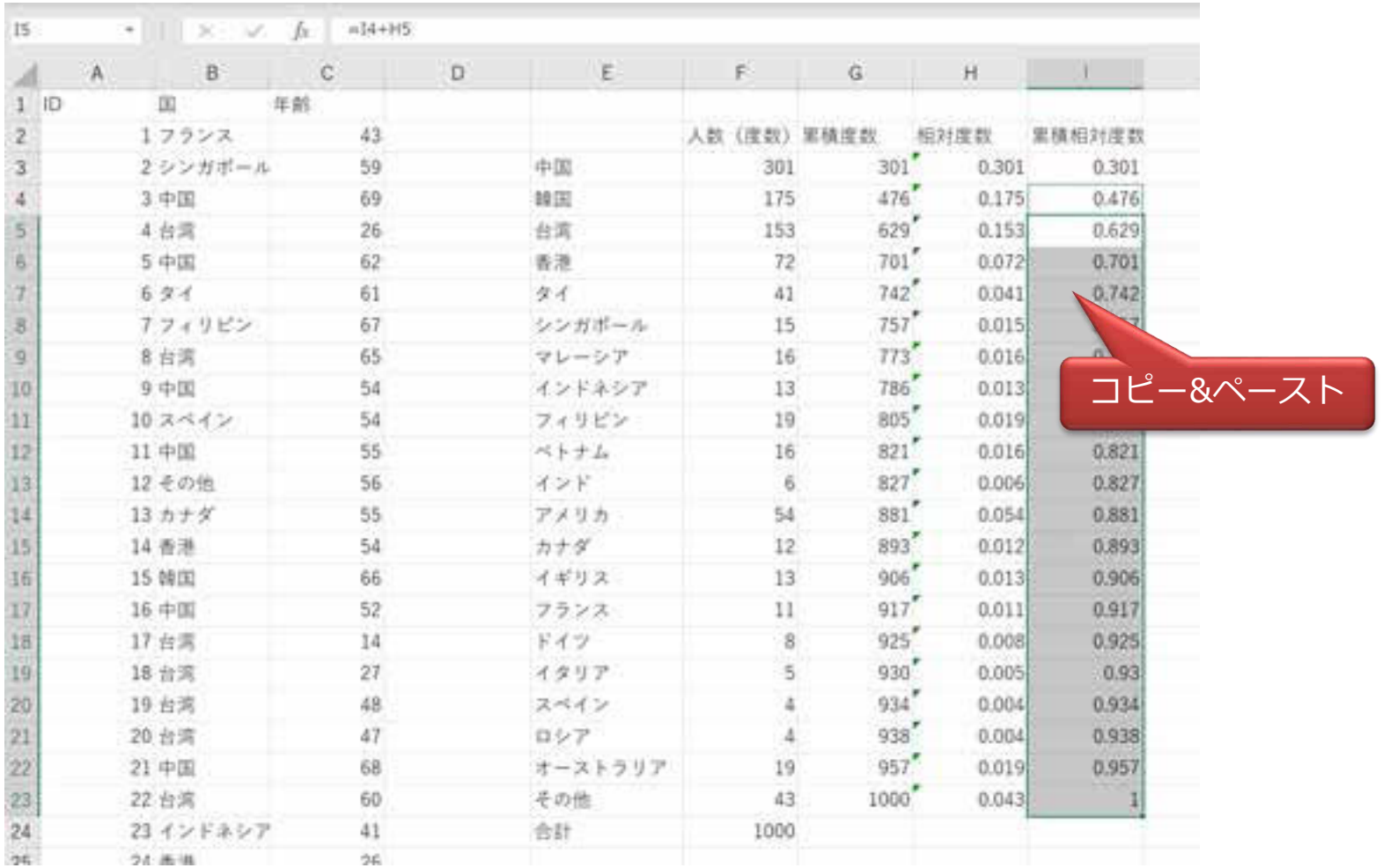
H4 =I3+H4

	A	B	C	D	E	F	G	H	I	J
1	ID	国	年齢							
2		1 フランス	43			人数（度数）	累積度数	相対度数	累積相対度数	
3		2 シンガポール	59		中国	301	301	0.301	0.301	
4		3 中国	69		韓国	175	476	0.175	=I3+H4	
5		4 台湾	26		台湾	153	629	0.153		
6		5 中国	62		香港	72	701	0.072		
7		6 タイ	61		タイ	41	742			
8		7 フィリピン	67		シンガポール	15	757			
9		8 台湾	65		マレーシア	16	773			
10		9 中国	54		インドネシア	13	786			
11		10 スペイン	54		フィリピン	19	805	0.019		
12		11 中国	55		ベトナム	16	821	0.016		
13		12 その他	56		インド	6	827	0.006		
14		13 カナダ	55		アメリカ	54	881	0.054		
15		14 香港	54		カナダ	12	893	0.012		
16		15 韓国	66		イギリス	13	906	0.013		
17		16 中国	52		フランス	11	917	0.011		
18		17 台湾	14		ドイツ	8	925	0.008		
19		18 台湾	27		イタリア	5	930	0.005		
20		19 台湾	48		スペイン	4	934	0.004		
21		20 台湾	47		ロシア	4	938	0.004		
22		21 中国	68		オーストラリア	19	957	0.019		
23		22 台湾	60		その他	43	1000	0.043		
24		23 インドネシア	41		合計	1000				

2行目以降は前の行の合計と相対度数を足し合わせる。

ハンズオン（14）

- 度数分布表の完成

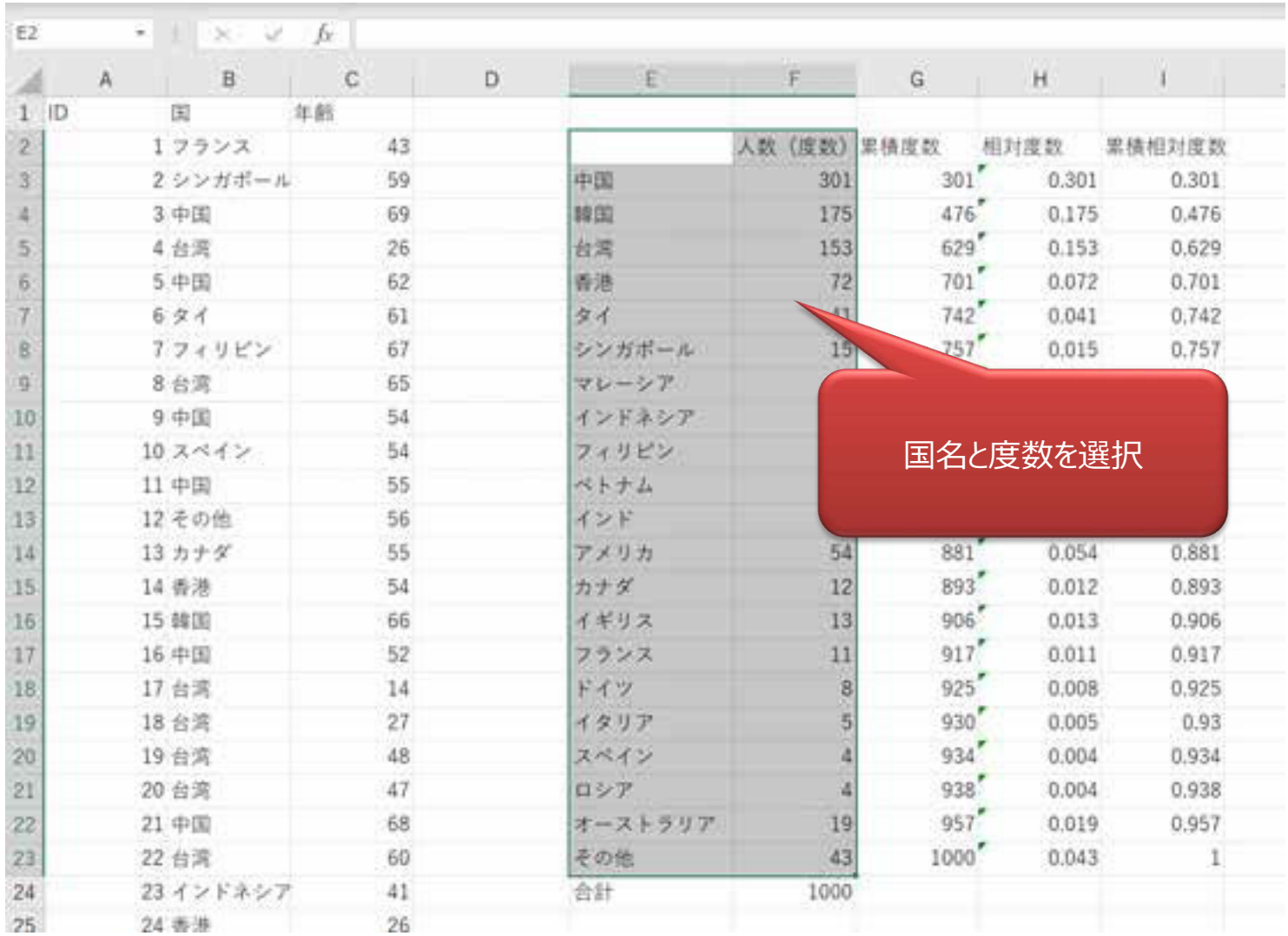


I5 =I4+H5

	A	B	C	D	E	F	G	H	I
1	ID	国	年齢						
2		1 フランス	43			人数（度数）	累積度数	相対度数	累積相対度数
3		2 シンガポール	59		中国	301	301	0.301	0.301
4		3 中国	69		韓国	175	476	0.175	0.476
5		4 台湾	26		台湾	153	629	0.153	0.629
6		5 中国	62		香港	72	701	0.072	0.701
7		6 タイ	61		タイ	41	742	0.041	0.742
8		7 フィリピン	67		シンガポール	15	757	0.015	
9		8 台湾	65		マレーシア	16	773	0.016	
10		9 中国	54		インドネシア	13	786	0.013	
11		10 スペイン	54		フィリピン	19	805	0.019	
12		11 中国	55		ベトナム	16	821	0.016	0.821
13		12 その他	56		インド	6	827	0.006	0.827
14		13 カナダ	55		アメリカ	54	881	0.054	0.881
15		14 香港	54		カナダ	12	893	0.012	0.893
16		15 韓国	66		イギリス	13	906	0.013	0.906
17		16 中国	52		フランス	11	917	0.011	0.917
18		17 台湾	14		ドイツ	8	925	0.008	0.925
19		18 台湾	27		イタリア	5	930	0.005	0.93
20		19 台湾	48		スペイン	4	934	0.004	0.934
21		20 台湾	47		ロシア	4	938	0.004	0.938
22		21 中国	68		オーストラリア	19	957	0.019	0.957
23		22 台湾	60		その他	43	1000	0.043	1
24		23 インドネシア	41		合計	1000			
25		24 香港	26						

コピー&ペースト

ハンズオン（15）

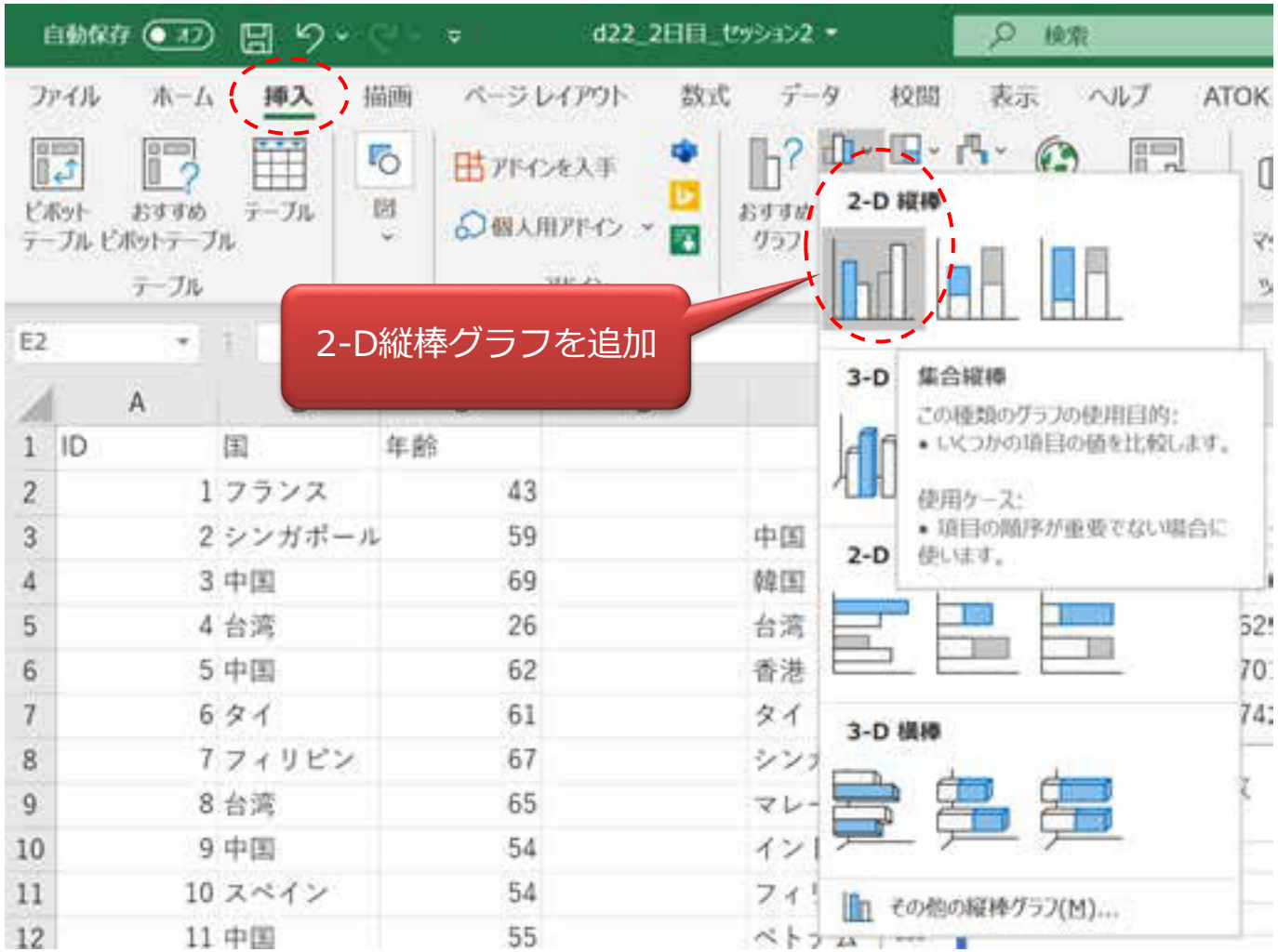


E2

	A	B	C	D	E	F	G	H	I
1	ID	国	年齢						
2	1	フランス	43			人数（度数）	累積度数	相対度数	累積相対度数
3	2	シンガポール	59		中国	301	301	0.301	0.301
4	3	中国	69		韓国	175	476	0.175	0.476
5	4	台湾	26		台湾	153	629	0.153	0.629
6	5	中国	62		香港	72	701	0.072	0.701
7	6	タイ	61		タイ	41	742	0.041	0.742
8	7	フィリピン	67		シンガポール	15	757	0.015	0.757
9	8	台湾	65		マレーシア				
10	9	中国	54		インドネシア				
11	10	スペイン	54		フィリピン				
12	11	中国	55		ベトナム				
13	12	その他	56		インド				
14	13	カナダ	55		アメリカ	54	881	0.054	0.881
15	14	香港	54		カナダ	12	893	0.012	0.893
16	15	韓国	66		イギリス	13	906	0.013	0.906
17	16	中国	52		フランス	11	917	0.011	0.917
18	17	台湾	14		ドイツ	8	925	0.008	0.925
19	18	台湾	27		イタリア	5	930	0.005	0.93
20	19	台湾	48		スペイン	4	934	0.004	0.934
21	20	台湾	47		ロシア	4	938	0.004	0.938
22	21	中国	68		オーストラリア	19	957	0.019	0.957
23	22	台湾	60		その他	43	1000	0.043	1
24	23	インドネシア	41		合計	1000			
25	24	香港	26						

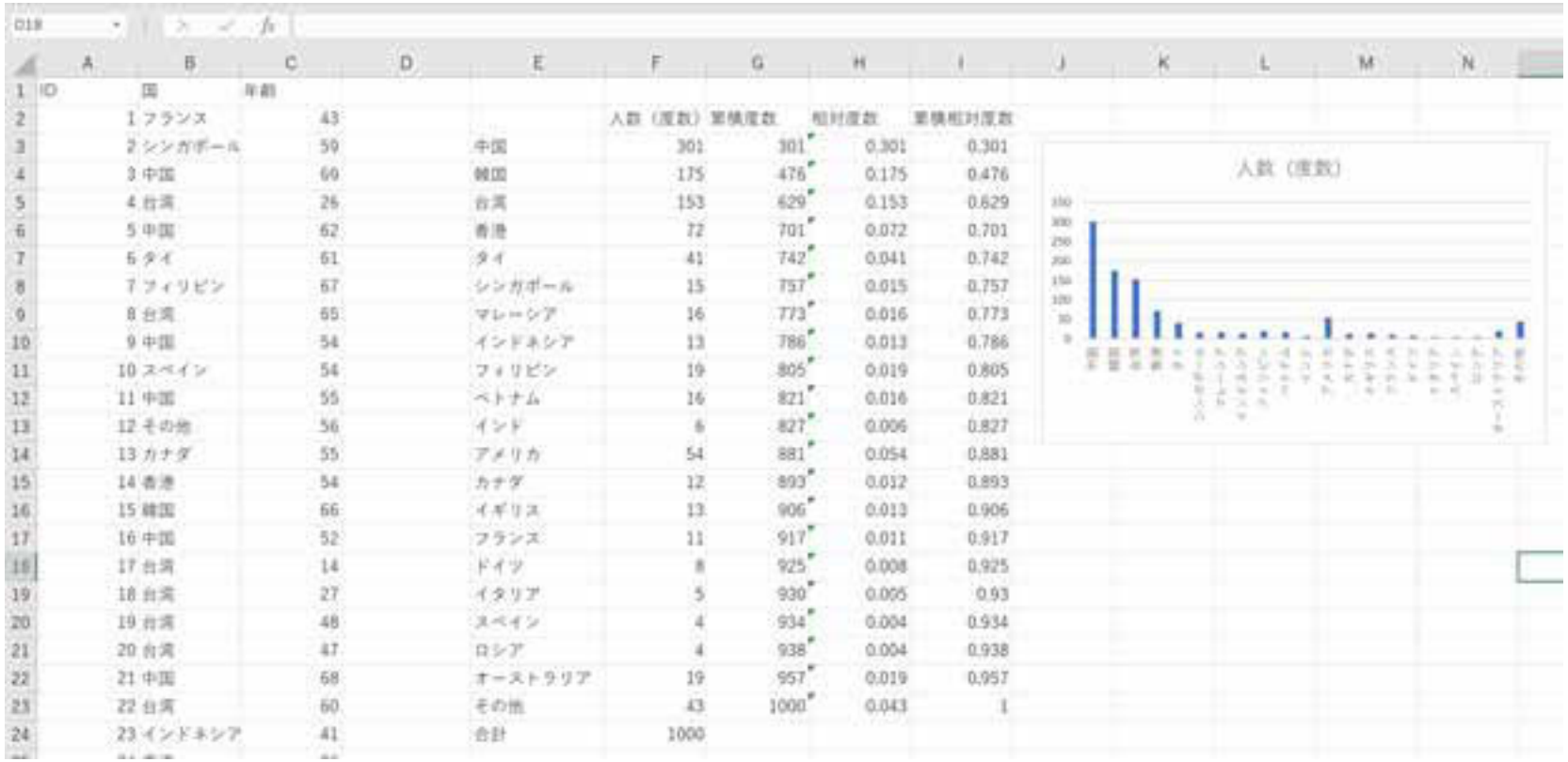
国名と度数を選択

ハンズオン（16）


自動保存
d22_2日目_セッション2
検索
ファイル
ホーム
挿入
描画
ページレイアウト
数式
データ
校閲
表示
ヘルプ
ATOK
ピボットテーブル
おすすめピボットテーブル
テーブル
テーブル
図
アドインを入手
個人用アドイン
おすすめグラフ
2-D 縦棒
2-D縦棒グラフを追加
3-D
集合縦棒
この種類のグラフの使用目的:
• いくつかの項目の値を比較します。
使用ケース:
• 項目の順序が重要でない場合に使います。
2-D
3-D 横棒
その他の縦棒グラフ(M)...
E2
A
1 ID 国 年齢
2 1 フランス 43
3 2 シンガポール 59
4 3 中国 69
5 4 台湾 26
6 5 中国 62
7 6 タイ 61
8 7 フィリピン 67
9 8 台湾 65
10 9 中国 54
11 10 スペイン 54
12 11 中国 55
中国
韓国
台湾
香港
タイ

ハンズオン（17）

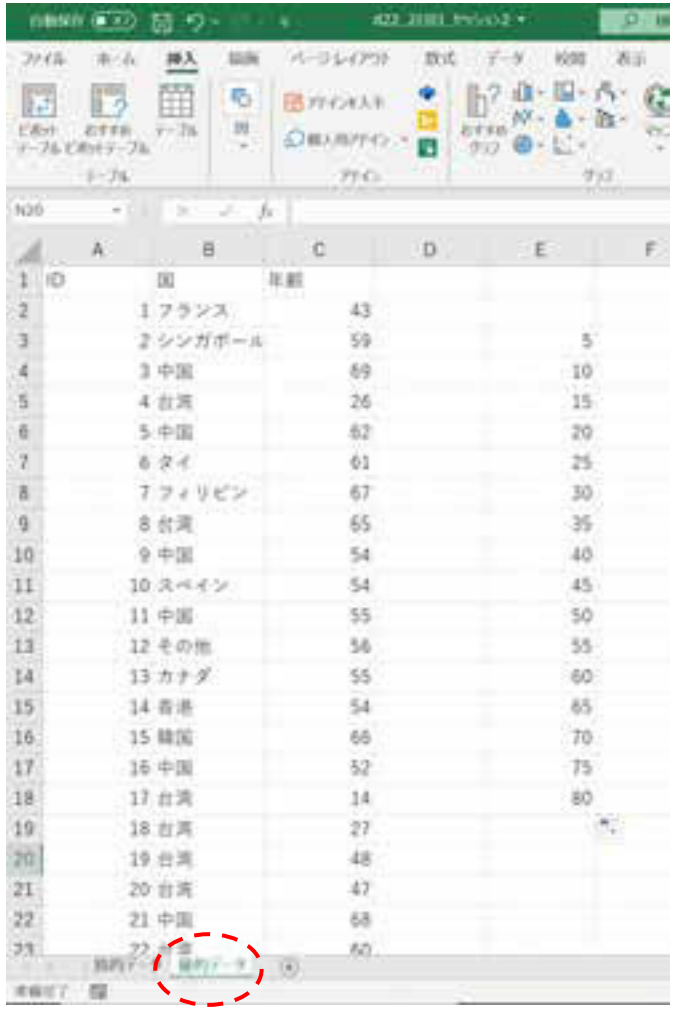
- 質的データのグラフが完成



	A	B	C	D	E	F	G	H	I
1	ID	国	年齢						
2		1 フランス	43			人数（度数）	累積度数	相対度数	累積相対度数
3		2 シンガポール	59		中国	301	301	0.301	0.301
4		3 中国	69		韓国	175	476	0.175	0.476
5		4 台湾	26		台湾	153	629	0.153	0.629
6		5 中国	62		香港	72	701	0.072	0.701
7		6 タイ	61		タイ	41	742	0.041	0.742
8		7 フィリピン	67		シンガポール	15	757	0.015	0.757
9		8 台湾	65		マレーシア	16	773	0.016	0.773
10		9 中国	54		インドネシア	13	786	0.013	0.786
11		10 スペイン	54		フィリピン	19	805	0.019	0.805
12		11 中国	55		ベトナム	16	821	0.016	0.821
13		12 その他	56		インド	6	827	0.006	0.827
14		13 カナダ	55		アメリカ	54	881	0.054	0.881
15		14 香港	54		カナダ	12	893	0.012	0.893
16		15 韓国	66		イギリス	13	906	0.013	0.906
17		16 中国	52		フランス	11	917	0.011	0.917
18		17 台湾	14		ドイツ	8	925	0.008	0.925
19		18 台湾	27		イタリア	5	930	0.005	0.93
20		19 台湾	48		スペイン	4	934	0.004	0.934
21		20 台湾	47		ロシア	4	938	0.004	0.938
22		21 中国	68		オーストラリア	19	957	0.019	0.957
23		22 台湾	60		その他	43	1000	0.043	1
24		23 インドネシア	41		合計	1000			

ハンズオン（18）

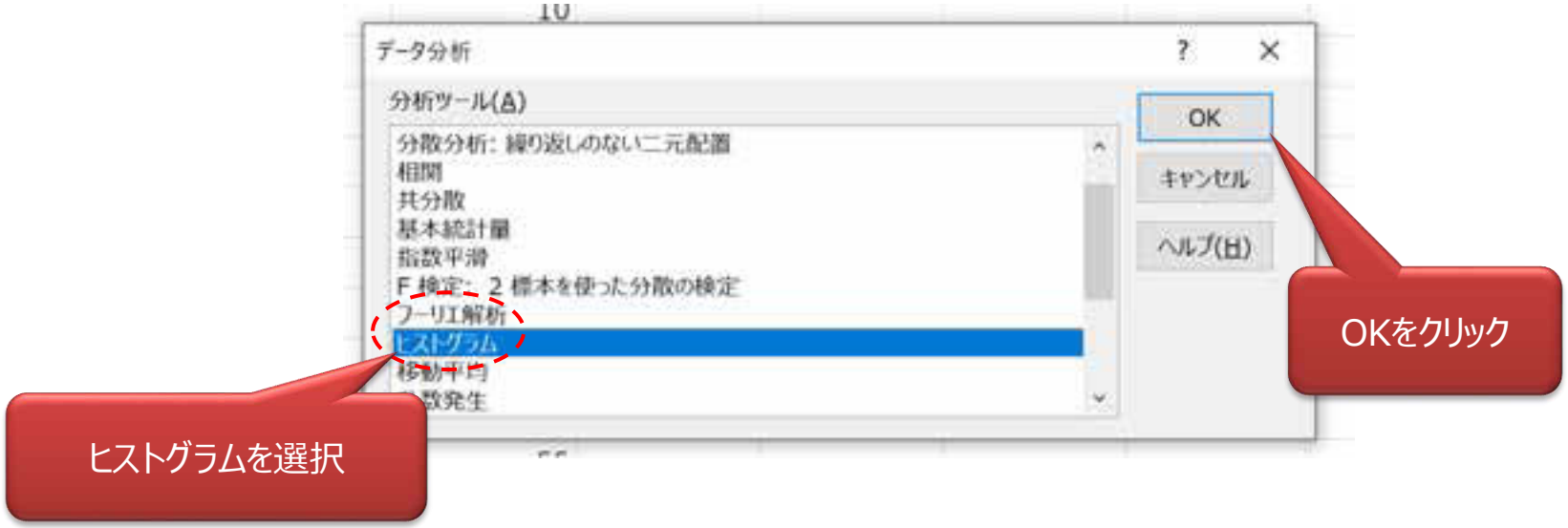
- シート「量的データ」



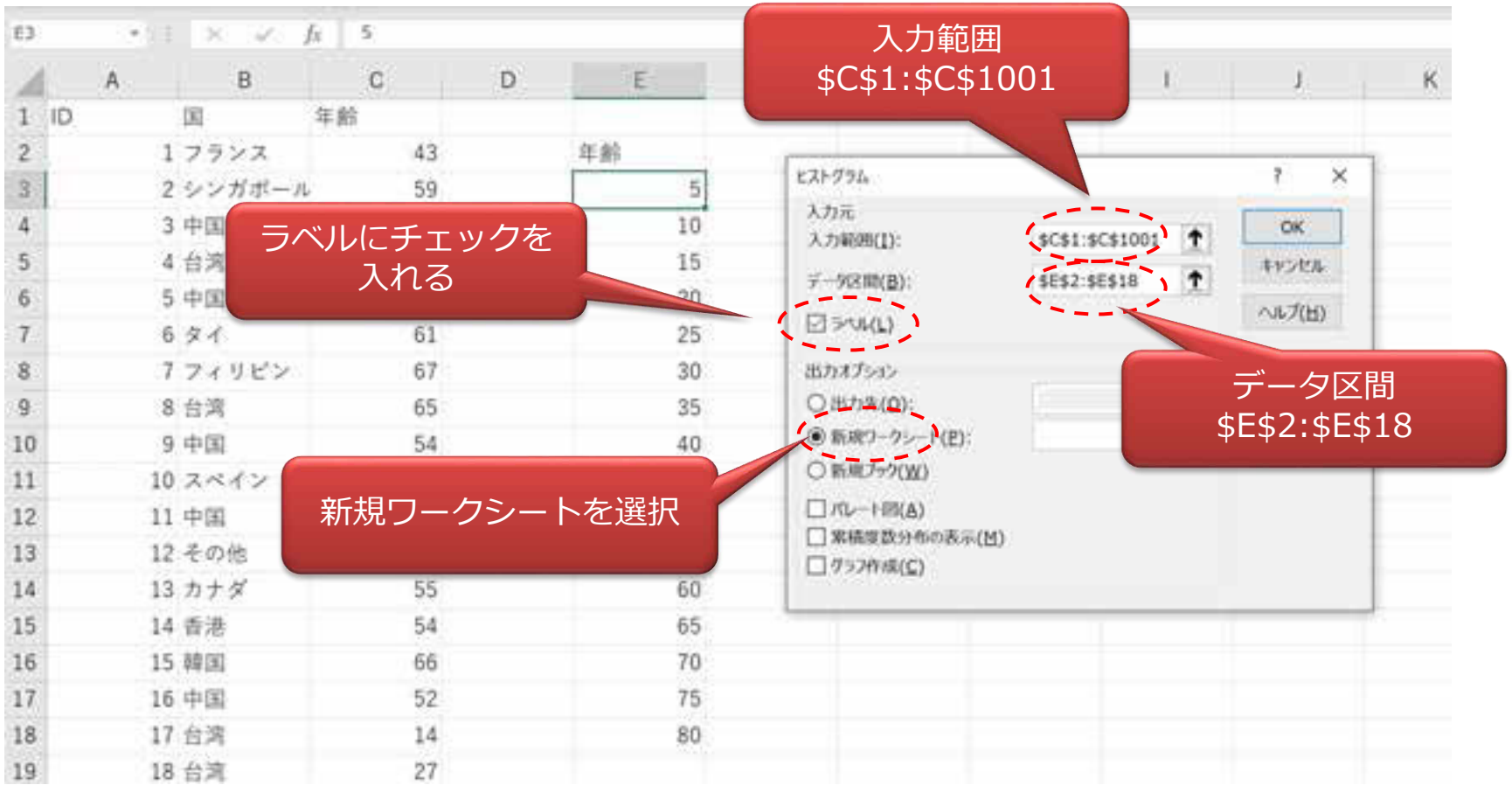
	A	B	C	D	E	F
1	ID	国	年齢			
2		1 フランス	43			
3		2 シンガポール	59		5	
4		3 中国	69		10	
5		4 台湾	26		15	
6		5 中国	62		20	
7		6 タイ	61		25	
8		7 フィリピン	67		30	
9		8 台湾	65		35	
10		9 中国	54		40	
11		10 スペイン	54		45	
12		11 中国	55		50	
13		12 その他	56		55	
14		13 カナダ	55		60	
15		14 香港	54		65	
16		15 韓国	66		70	
17		16 中国	52		75	
18		17 台湾	14		80	
19		18 台湾	27			
20		19 台湾	48			
21		20 台湾	47			
22		21 中国	68			

ハンズオン（19）


データ分析をクリック

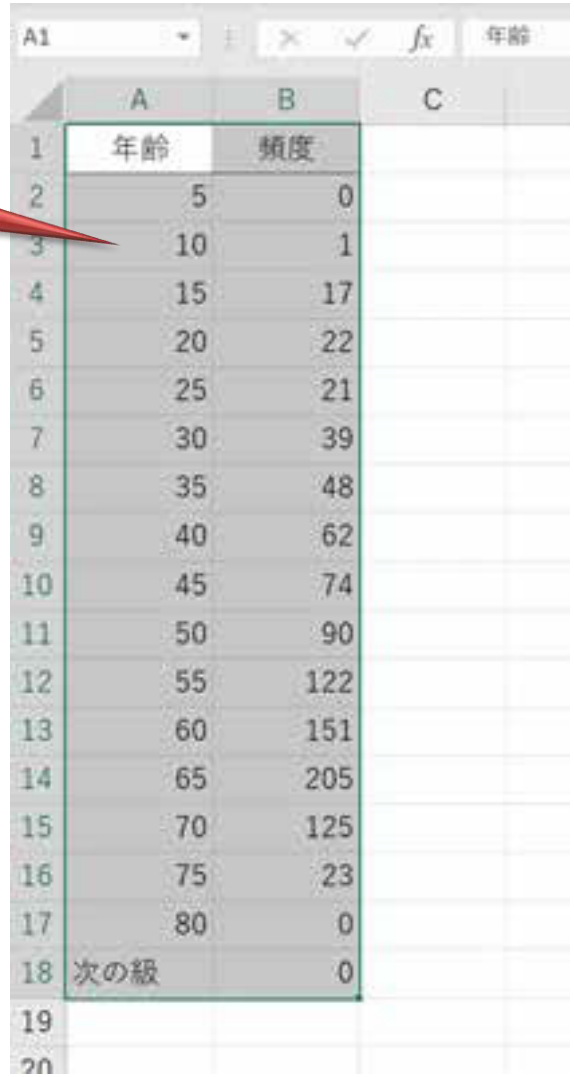

データ分析
分析ツール(A)
分散分析: 繰り返しのない二元配置
相関
共分散
基本統計量
指数平滑
F 検定: 2 標本を使った分散の検定
フーリエ解析
ヒストグラム
OK
キャンセル
ヘルプ(H)
ヒストグラムを選択
OKをクリック

ハンズオン（20）


入力範囲
$C$1:$C$1001
ラベルにチェックを
入れる
データ区間
$E$2:$E$18
新規ワークシートを選択
ヒストグラム
入力元
入力範囲(I):
$C$1:$C$1001
データ区間(B):
$E$2:$E$18
ラベル(L)
出力オプション
新規ワークシート(P):
新規ブック(W)
パレート図(A)
累積度数分布の表示(M)
グラフ作成(C)
OK
キャンセル
ヘルプ(H)

ハンズオン（21）

別のワークシートに度数集計結果が表示される。



	A	B	C
1	年齢	頻度	
2	5	0	
3	10	1	
4	15	17	
5	20	22	
6	25	21	
7	30	39	
8	35	48	
9	40	62	
10	45	74	
11	50	90	
12	55	122	
13	60	151	
14	65	205	
15	70	125	
16	75	23	
17	80	0	
18	次の級	0	
19			
20			

ハンズオン（22）

- 度数分布表の作成

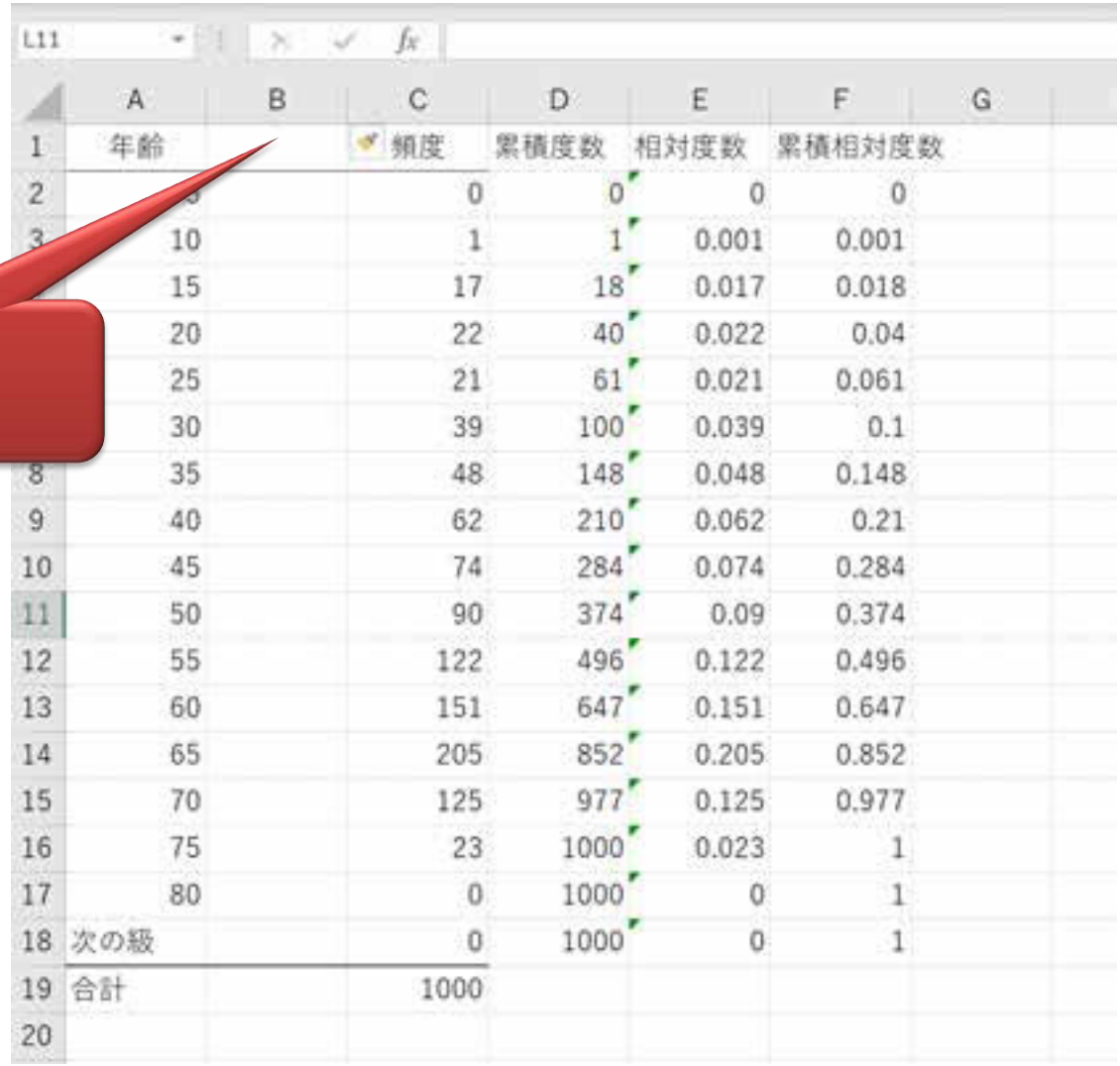
質的データと同様にして度数分布表を完成させましょう。

M10

	A	B	C	D	E	F
1	年齢	頻度	累積度数	相対度数	累積相対度数	
2	5	0	0	0	0	
3	10	1	1	0.001	0.001	
4	15	17	18	0.017	0.018	
5	20	22	40	0.022	0.04	
6	25	21	61	0.021	0.061	
7	30	39	100	0.039	0.1	
8	35	48	148	0.048	0.148	
9	40	62	210	0.062	0.21	
10	45	74	284	0.074	0.284	
11	50	90	374	0.09	0.374	
12	55	122	496	0.122	0.496	
13	60	151	647	0.151	0.647	
14	65	205	852	0.205	0.852	
15	70	125	977	0.125	0.977	
16	75	23	1000	0.023	1	
17	80	0	1000	0	1	
18	次の級	0	1000	0	1	
19	合計	1000				
20						

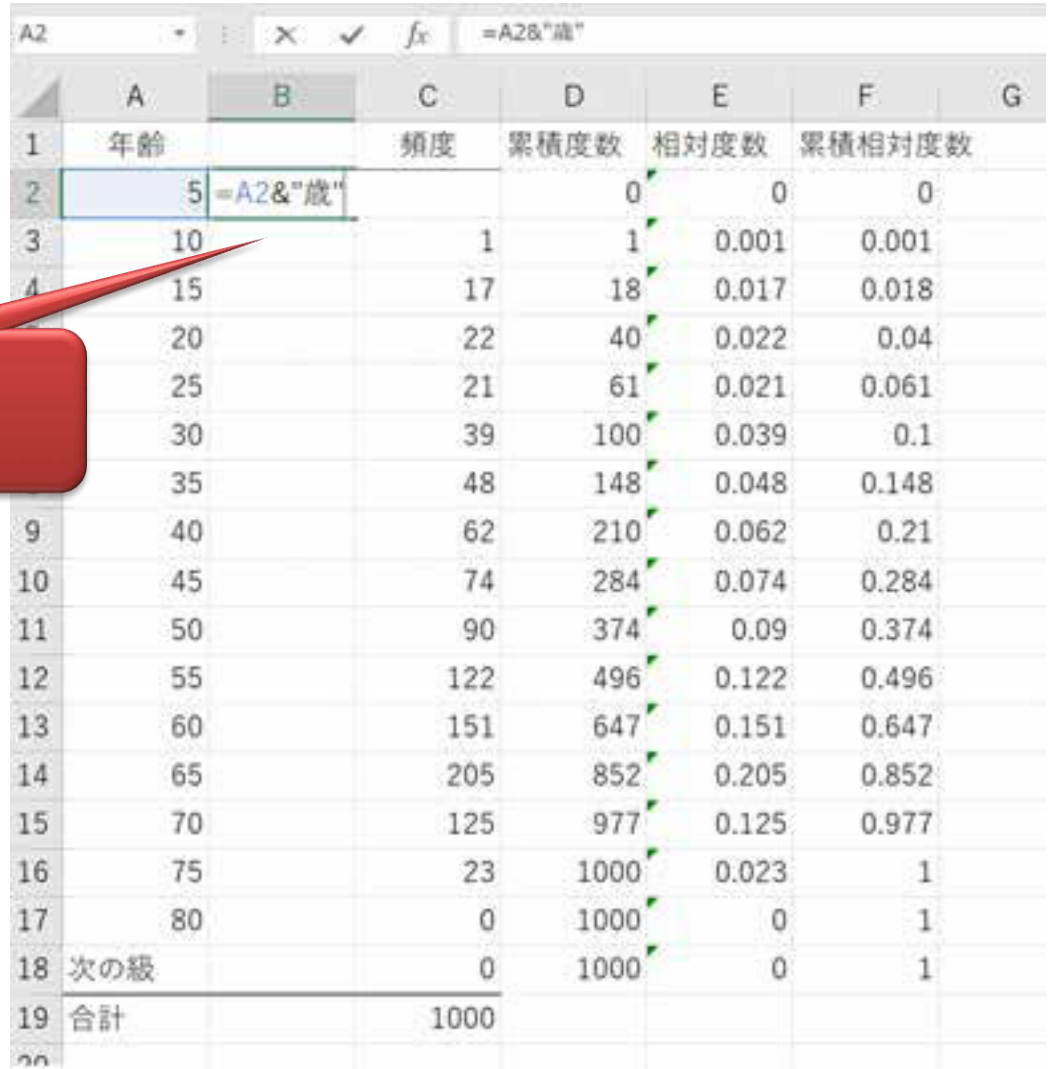
ハンズオン（23）

空白行を挿入



	A	B	C	D	E	F	G
1	年齢		頻度	累積度数	相対度数	累積相対度数	
2			0	0	0	0	
3	10		1	1	0.001	0.001	
4	15		17	18	0.017	0.018	
5	20		22	40	0.022	0.04	
6	25		21	61	0.021	0.061	
7	30		39	100	0.039	0.1	
8	35		48	148	0.048	0.148	
9	40		62	210	0.062	0.21	
10	45		74	284	0.074	0.284	
11	50		90	374	0.09	0.374	
12	55		122	496	0.122	0.496	
13	60		151	647	0.151	0.647	
14	65		205	852	0.205	0.852	
15	70		125	977	0.125	0.977	
16	75		23	1000	0.023	1	
17	80		0	1000	0	1	
18	次の級		0	1000	0	1	
19	合計		1000				
20							

ハンズオン（24）



	A	B	C	D	E	F	G
1	年齢		頻度	累積度数	相対度数	累積相対度数	
2	5	=A2&"歳"		0	0	0	
3	10		1	1	0.001	0.001	
4	15		17	18	0.017	0.018	
5	20		22	40	0.022	0.04	
6	25		21	61	0.021	0.061	
7	30		39	100	0.039	0.1	
8	35		48	148	0.048	0.148	
9	40		62	210	0.062	0.21	
10	45		74	284	0.074	0.284	
11	50		90	374	0.09	0.374	
12	55		122	496	0.122	0.496	
13	60		151	647	0.151	0.647	
14	65		205	852	0.205	0.852	
15	70		125	977	0.125	0.977	
16	75		23	1000	0.023	1	
17	80		0	1000	0	1	
18	次の級		0	1000	0	1	
19	合計		1000				

ハンズオン（25）

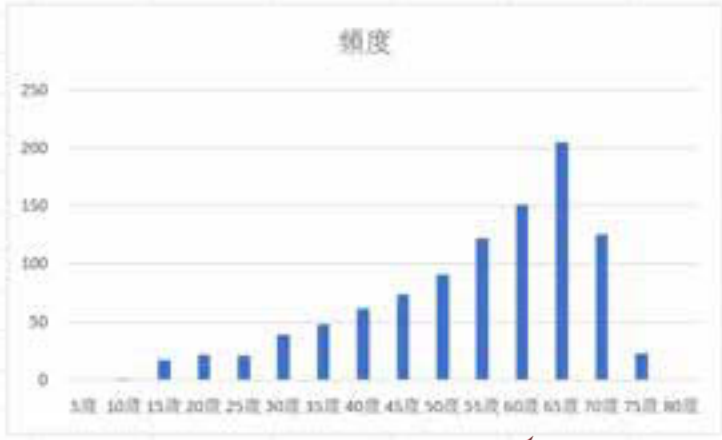


	A	B	C	D	E	F	G
1	年齢		頻度	累積度数	相対度数	累積相対度数	
2	5	5歳	0	0	0	0	
3	10	10歳	1	1	0.001	0.001	
4	15	15歳	17	18	0.017	0.018	
5	20	20歳	22	40	0.022	0.04	
6	25	25歳	21	61	0.021	0.061	
7	30	30歳	39	100	0.039	0.1	
8	35	35歳	48	148	0.048	0.148	
9	40	40歳	62	210	0.062	0.21	
10	45	45歳	74	284	0.074	0.284	
11	50	50歳			0.09	0.374	
12	55	55歳			0.122	0.496	
13	60	60歳	151	647	0.151	0.647	
14	65	65歳	205	852	0.205	0.852	
15	70	70歳	125	977	0.125	0.977	
16	75	75歳	23	1000	0.023	1	
17	80	80歳	0	1000	0	1	
18	次の級		0	1000	0	1	
19	合計		1000				
20							

ハンズオン（26）

- ヒストグラムの完成

	A	B	C	D	E	F
1	年齢		頻度	累積度数	相対度数	累積相対度数
2	5	5歳	0	0	0	0
3	10	10歳	1	1	0.001	0.001
4	15	15歳	17	18	0.017	0.018
5	20	20歳	22	40	0.022	0.04
6	25	25歳	21	61	0.021	0.061
7	30	30歳	39	100	0.039	0.1
8	35	35歳	48	148	0.048	0.148
9	40	40歳	62	210	0.062	0.21
10	45	45歳	74	284	0.074	0.284
11	50	50歳	90	374	0.09	0.374
12	55	55歳	122	496	0.122	0.496
13	60	60歳	151	647	0.151	0.647
14	65	65歳	205	852	0.205	0.852
15	70	70歳	125	977	0.125	0.977
16	75	75歳	23	1000	0.023	1
17	80	80歳	0	1000	0	1
18	次の級		0	1000	0	1
19	合計		1000			



質的データと同様にして
グラフを作成しましょう。

統計学の基本

統計学の基礎　⑥散布度の読み取りと意義

ワーク

標準偏差とヒストグラムの読み取り

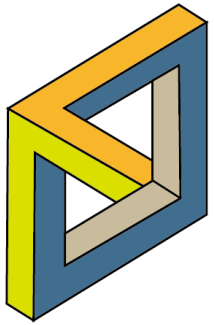
データから散布度を算出し、読み取ります。

再び、ファイル「演習データE1.xlsx」を開きましょう。

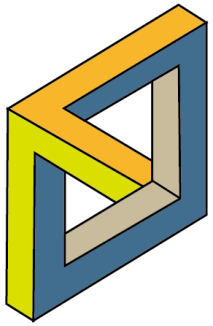
標準偏差を算出します。

ヒストグラムを作成します。

どのようなことが読み取れるか、考えてみましょう。



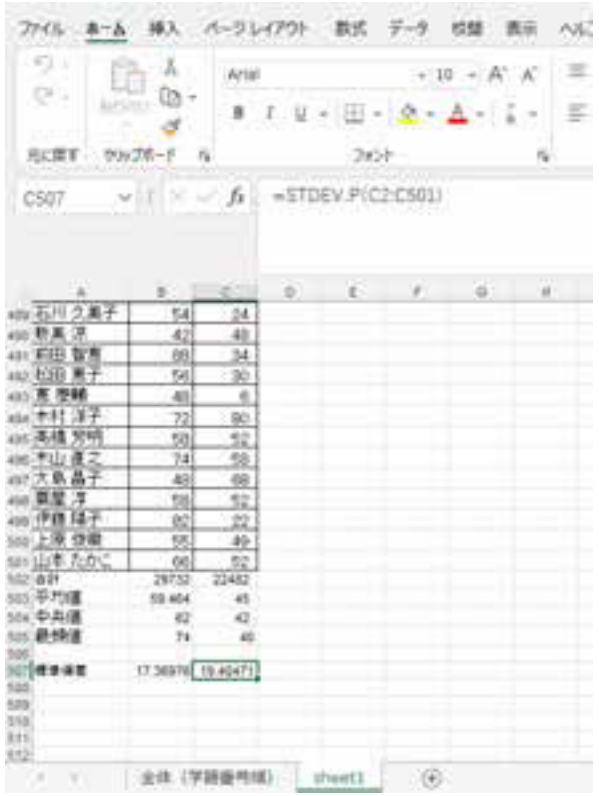
ハンズオン（標準偏差を1つの関数で）



標準偏差を求める

標準偏差は、1つの関数で算出する方法があります。



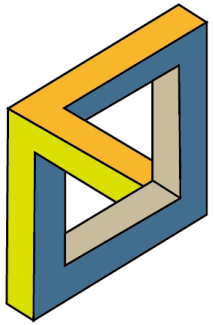


英語：17.4

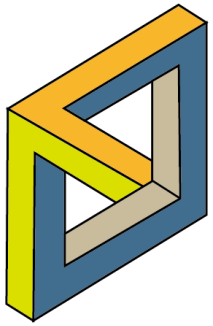
数学：19.4

偏差値とヒストグラムの読み取り

次にヒストグラムを作成しましょう。

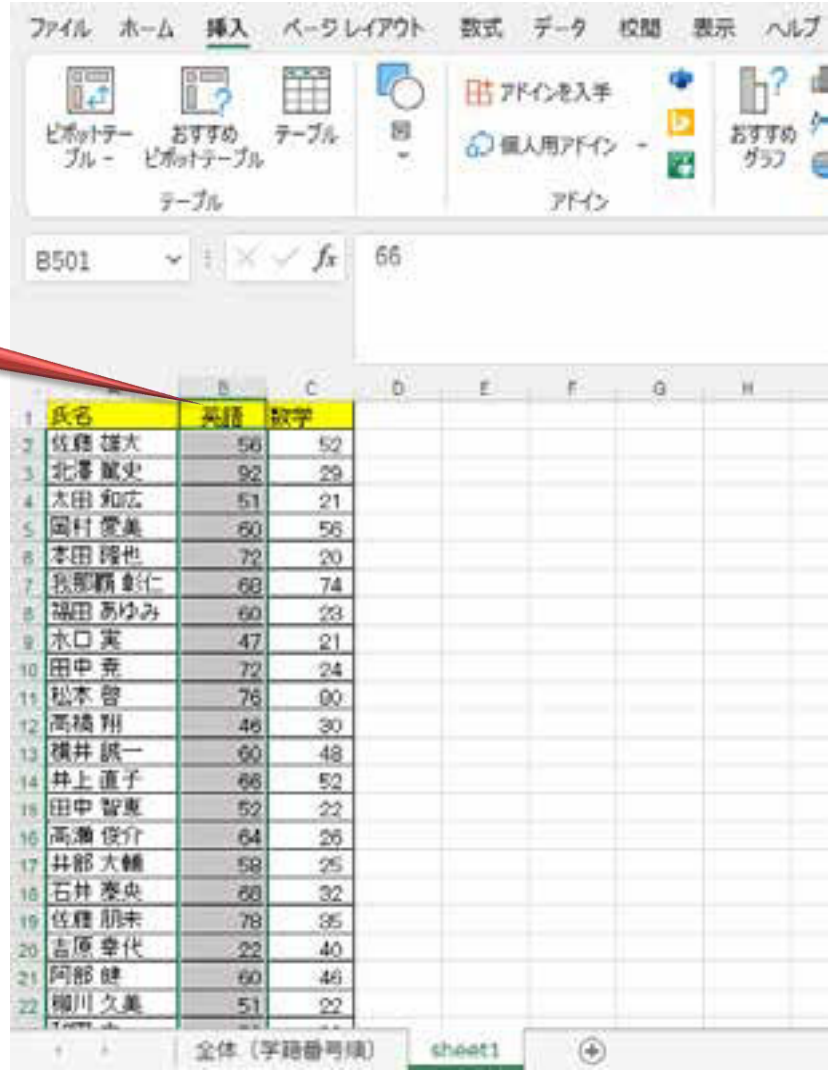


ハンズオン（答え合わせ）

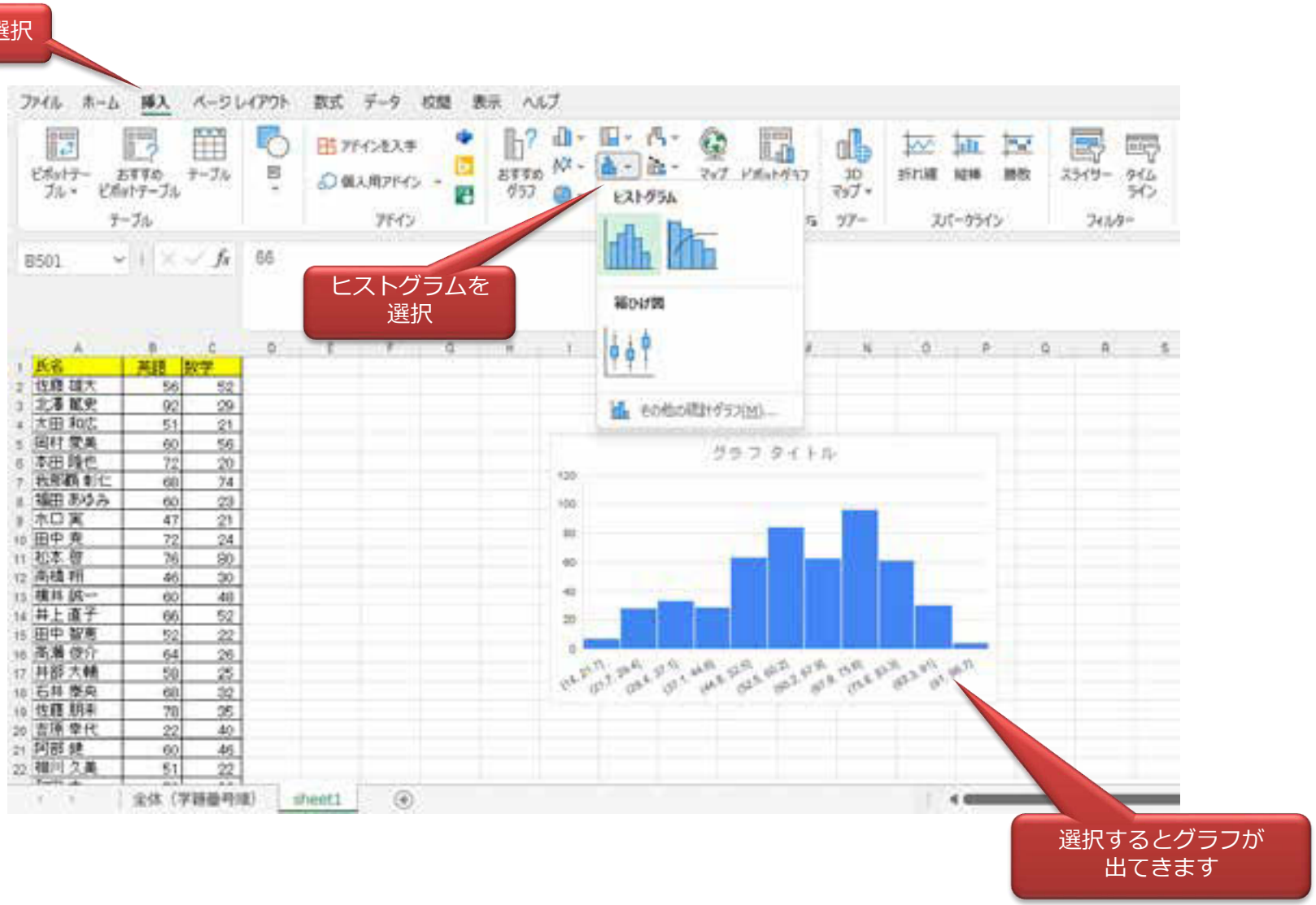


ハンズオン（1）

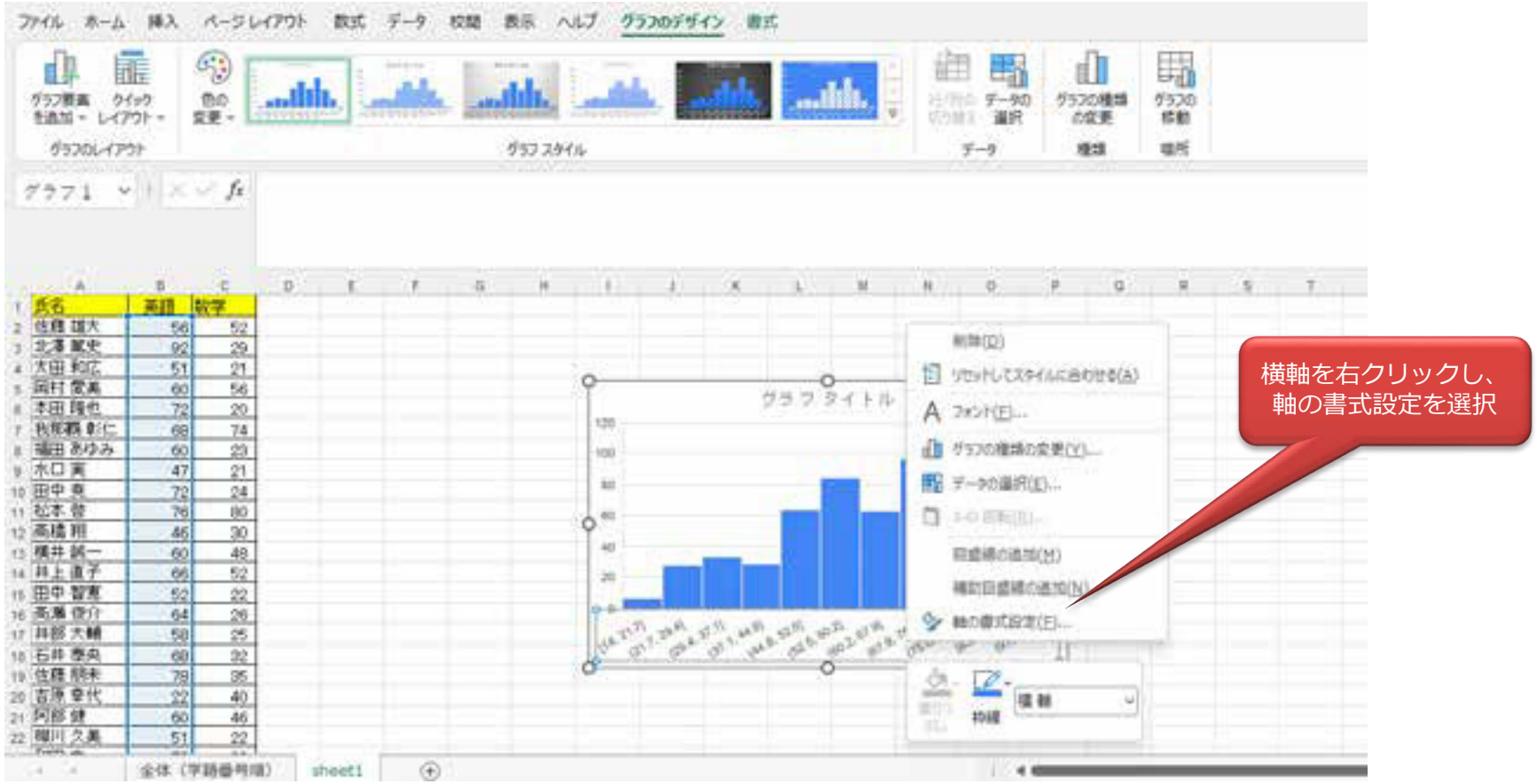
ヒストグラムを作る範囲を選択



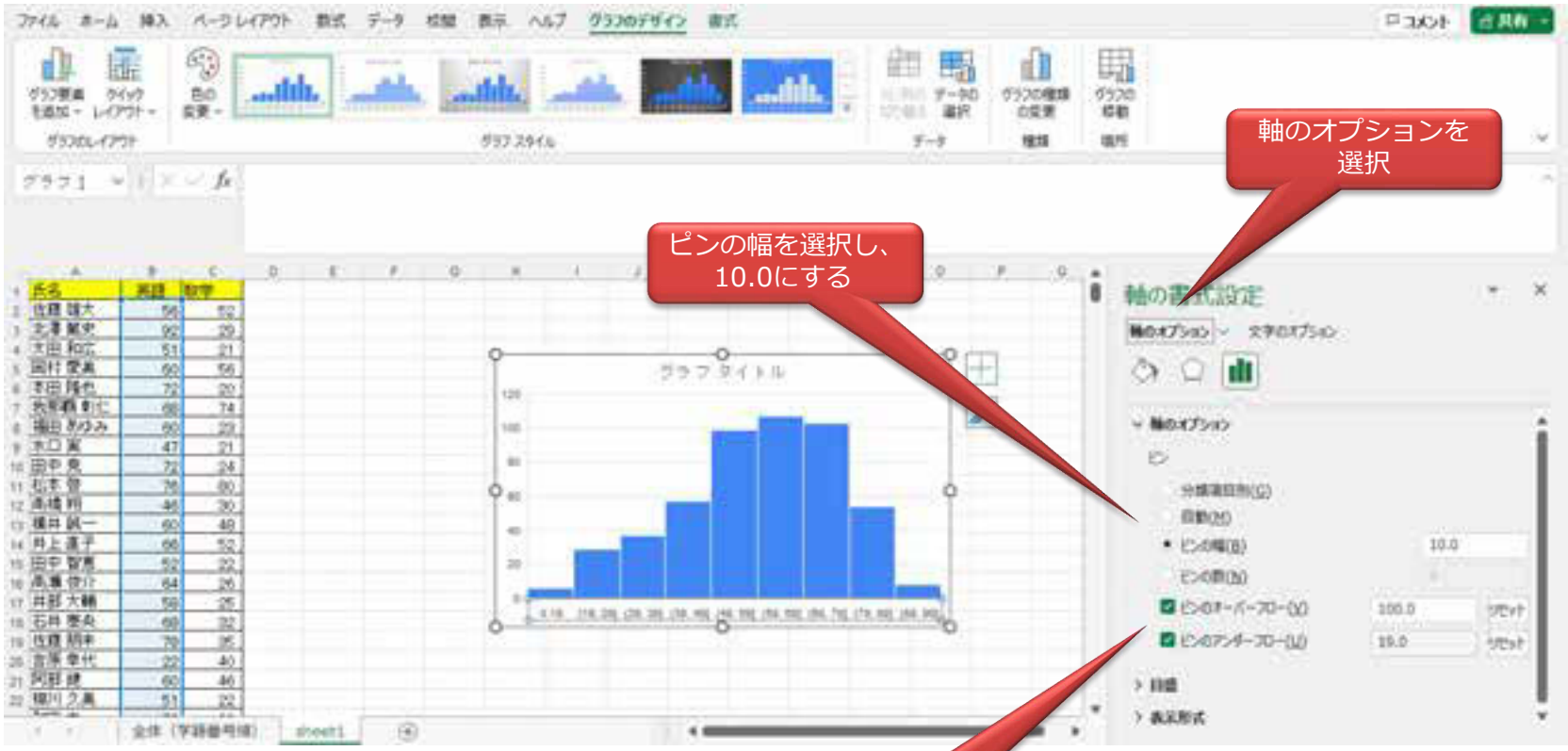
ハンズオン（2）



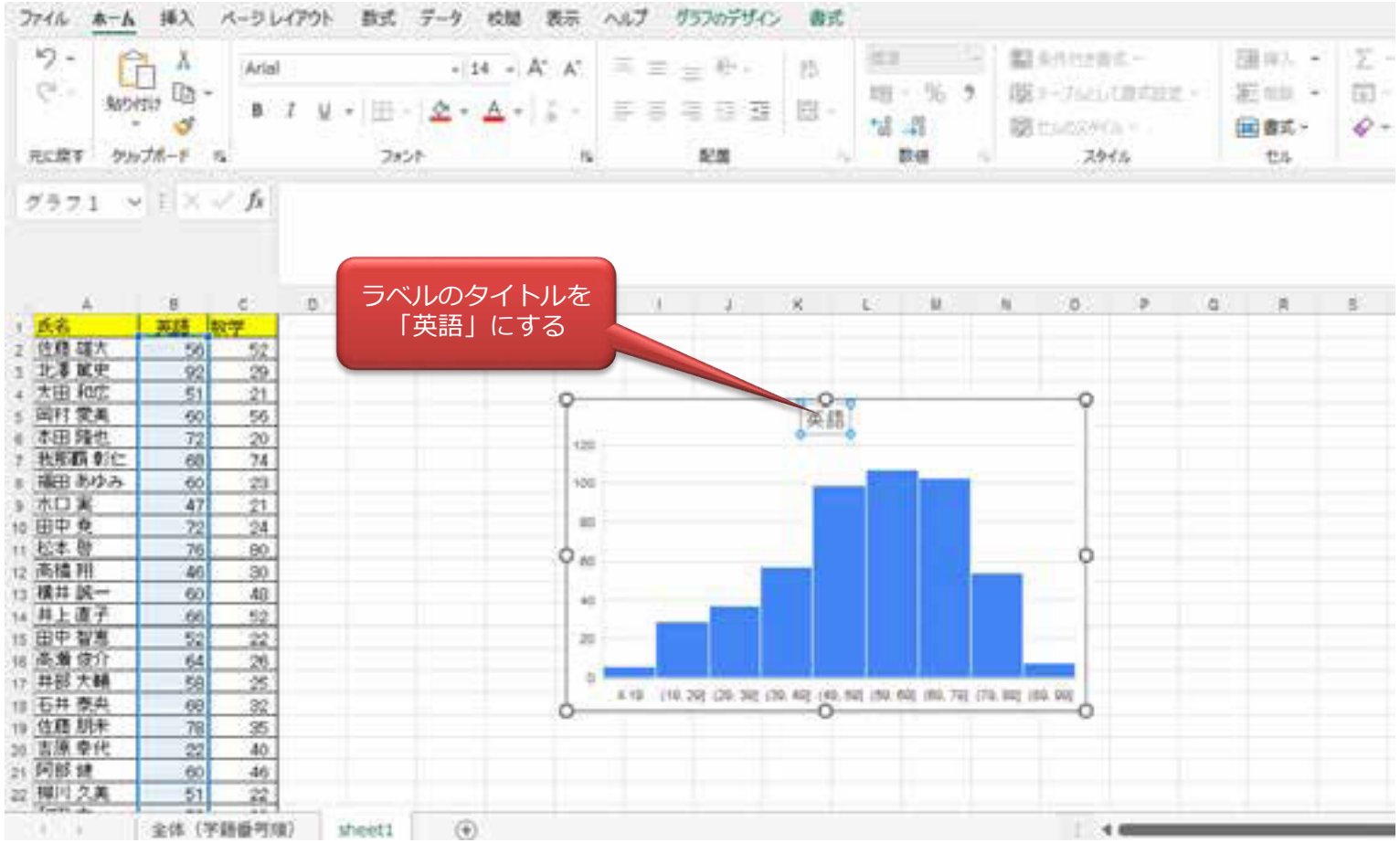
ハンズオン（3）



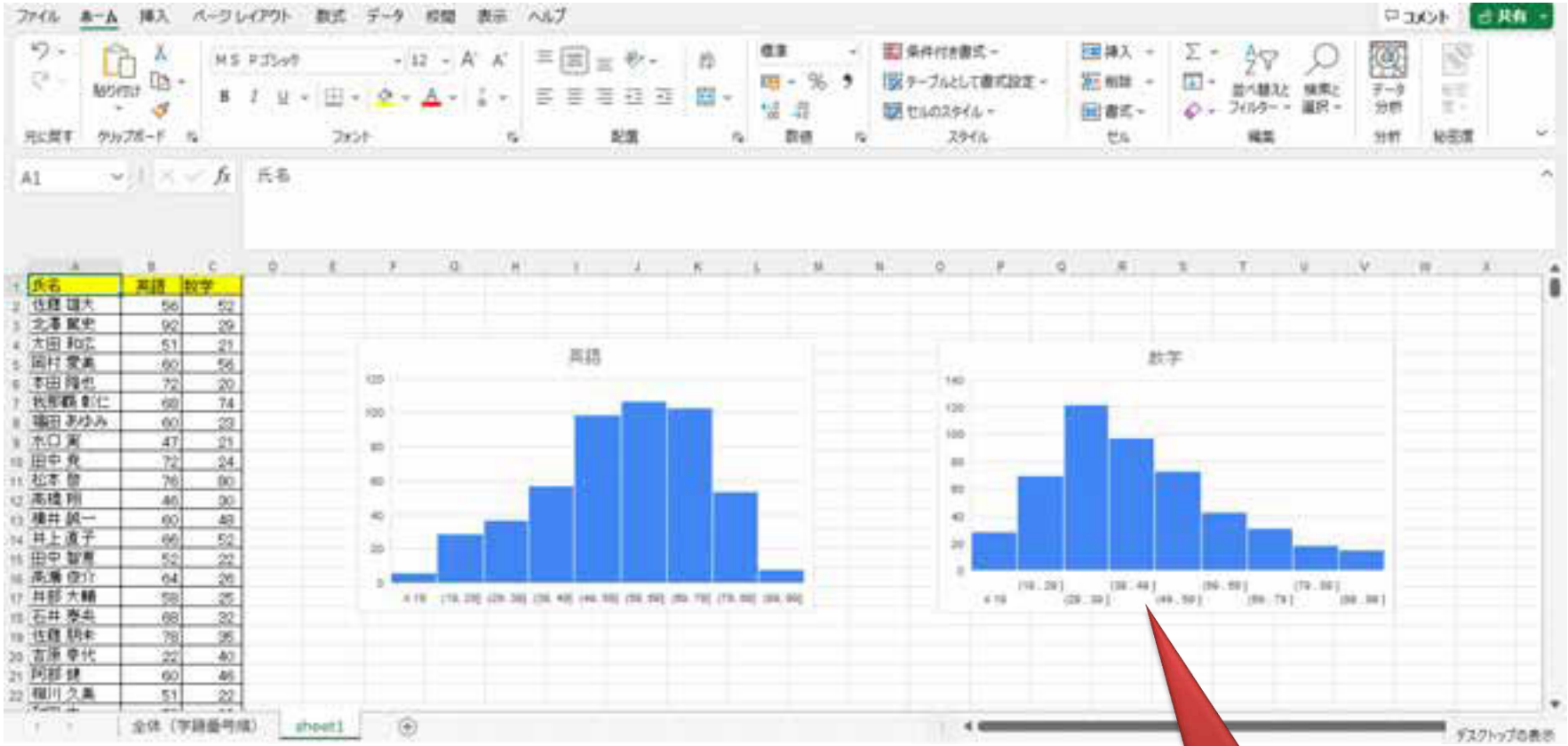
ハンズオン（4）


軸のオプションを選択
ピンの幅を選択し、10.0にする
オーバーフローを100.0に、アンダーフローを19.0にする。

ハンズオン（5）



ハンズオン（6）



ワーク

ワークの進め方（1）



後で見て分かるように
テーマを書いておきましょう

結果から読み取れそうなことを
グループで話し合いメモしましょう。

偏差値やヒストグラムから
どのようなことが読み取れるか
グループで考えてみましょう。

結論は出さなくて良いので
思いつくことをあげていきましょう。

代表値と散布度から見えること

代表値と標準偏差からだけでも
どのあたりにどれくらいの数値があるのかは分かります。
そこまで分かれば、そのデータの外観は掴めています。

この他にも様々な統計量の算出を覚えると、例えば
「2つのデータはどのうように関わっているか？」
というようなことも調べられるようになります。

