はじめての
統計データ利活用セミナー

日時：令和4年3月2日

講師：川原祐哉（和から株式会社）

講義スケジュール

章	内容	目標
第1章	統計学とは何か？	集計と可視化によるデータの特徴を把握する重要性を理解する
第2章	情報収集とデータ把握	ピボットテーブルの使い方、オープンデータのDL方法～分析方法を理解する
第3章	データ分析に必要な思考力	データを用いた意思決定の手順について理解する

講義スケジュール

章	内容	目標
第1章	統計学とは何か？	集計と可視化によるデータの特徴を把握する重要性を理解する
第2章	情報収集とデータ把握	ピボットテーブルの使い方、オープンデータのDL方法～分析方法を理解する
第3章	データ分析に必要な思考力	データを用いた意思決定の手順について理解する

第1章

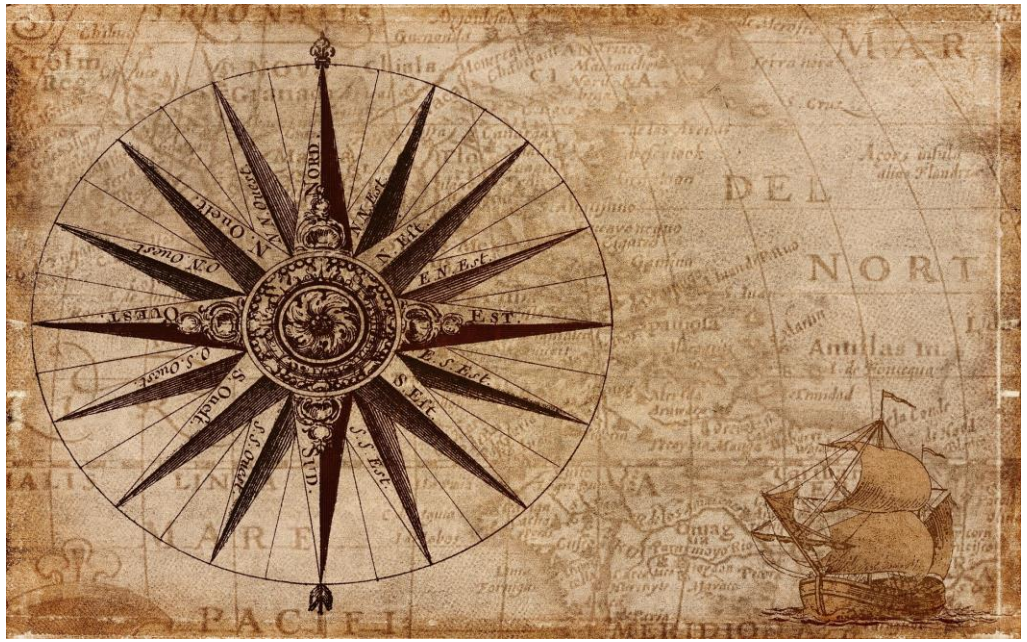
統計学とは何か？



統計学の歴史

歴史が重要な理由

学問







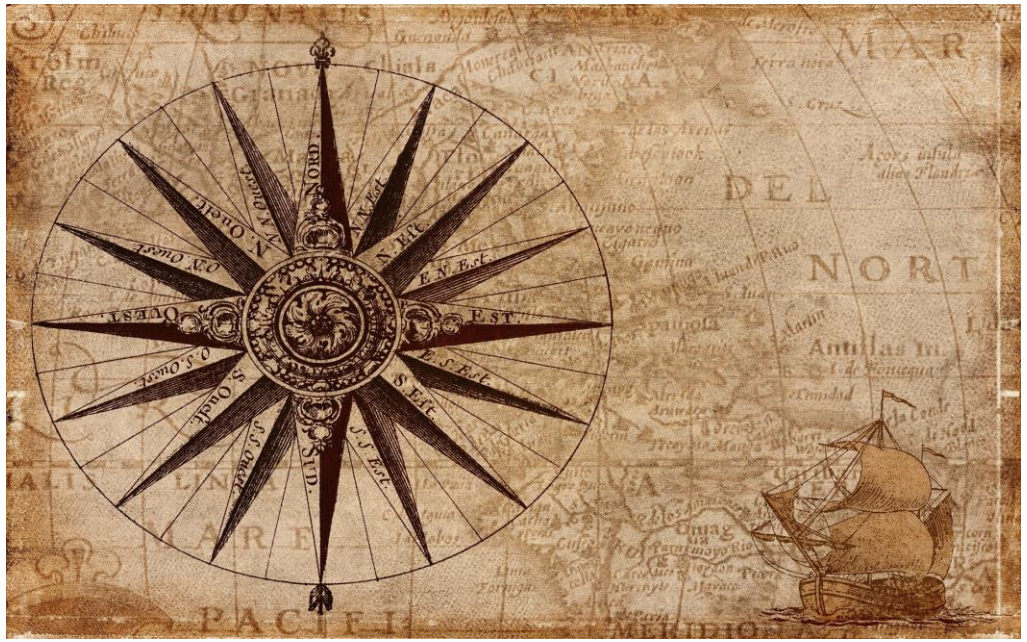


歴史が重要な理由

学問

必要性があってできたもの

・なぜ必要になったか？
・どんな背景があったか？　などなど…







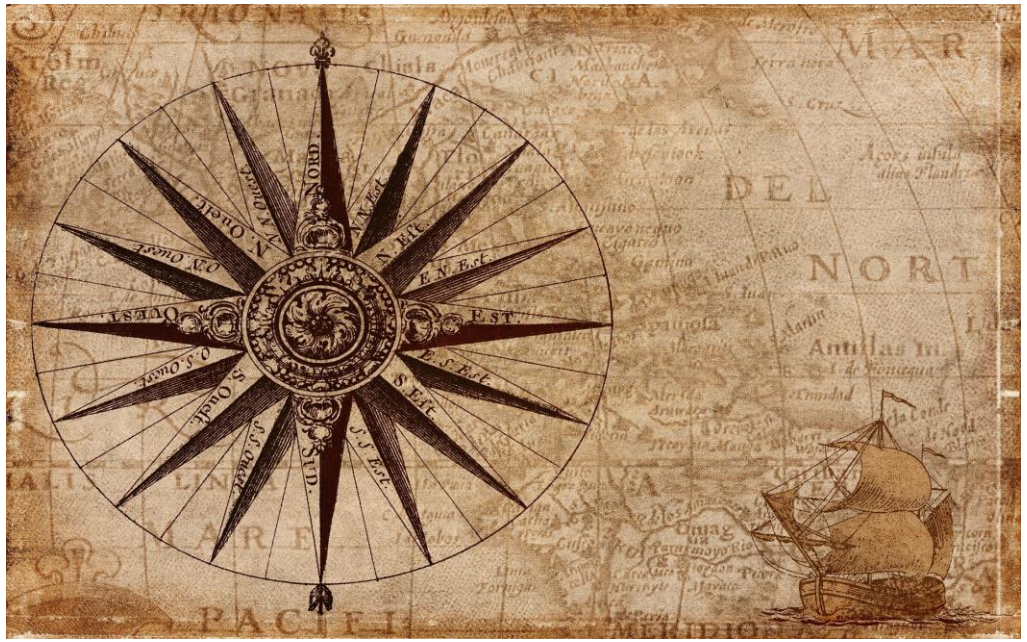


歴史が重要な理由

学問 → 必要性があってできたもの

・なぜ必要になったか？
・どんな背景があったか？　などなど…

→ 歴史









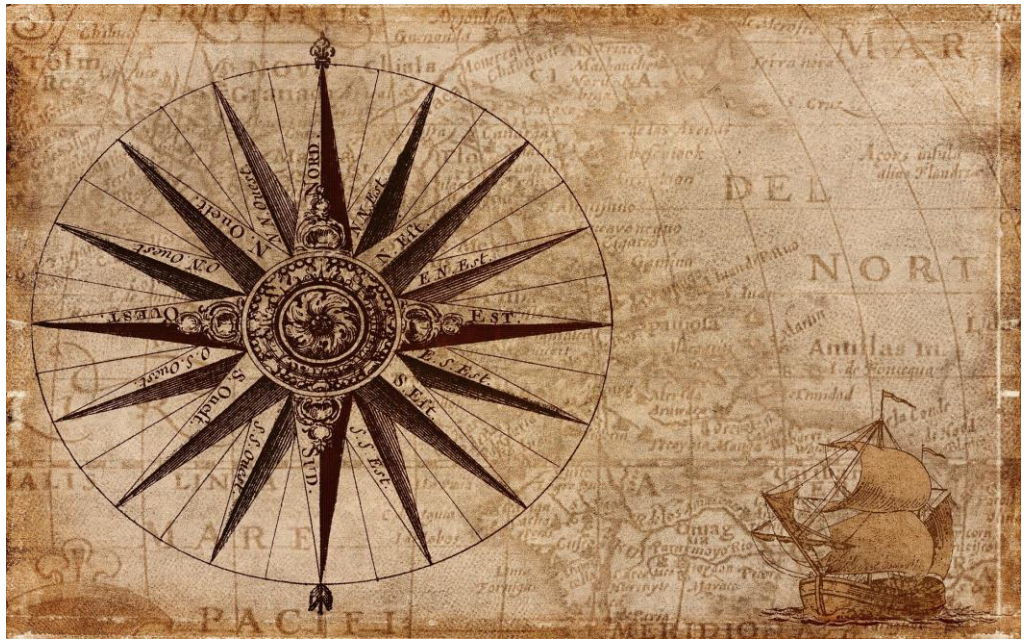
歴史が重要な理由

学問 → 必要性があってできたもの

・なぜ必要になったか？
・どんな背景があったか？　などなど…

→ 歴史

・モチベーションの向上
・楽しんで学ぶことができる









現代統計学の大きな3つの源流

1. 国の実態を捉えるための「統計」
2. 確率的事象を捉えるための「統計」
3. 大量の事象を捉えるための「統計」



「統計学」の語源

統計学

「統計学」の語源

統計学

Statistics

「統計学」の語源

統計学

Statistics

ラテン語「Status」

「統計学」の語源

統計学

Statistics

ラテン語「Status」



United “States”

「統計学」の語源

統計学

Statistics

ラテン語「Status」

国家・状態



United “States”

「統計学」の語源

統計学

Statistics

ラテン語「Status」

国家・状態

“国勢調査”



United “States”

「統計学」の語源

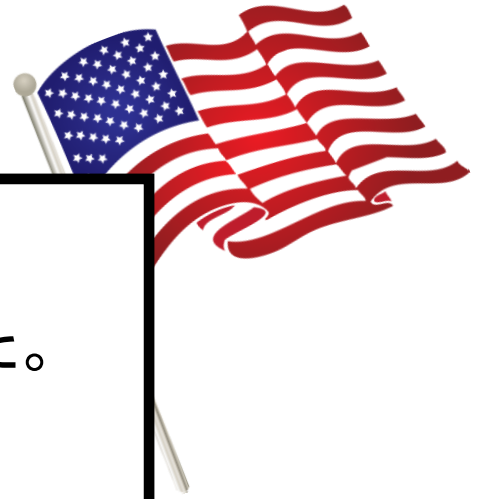
統計学

Statistics

国家・状態

“国勢調査”

統計学の始まりは**「国の実態を捉えるため」**だった。



United “States”

現代統計学の大きな3つの源流

1. 国の実態を捉えるための「統計」
2. 確率的事象を捉えるための「統計」
3. 大量の事象を捉えるための「統計」



現代統計学の大きな3つの源流

1. 国の実態を捉えるための「統計」
2. 確率的事象を捉えるための「統計」
3. 大量の事象を捉えるための「統計」



1．国の実態を捉えるための「統計」

BC3000年　BC100年　646年　1500年　1801年

古代　飛鳥～平安　ルネサンス　産業革命

ピラミッド建設のために人口調査
バビロニアで需給バランスの記録

ローマ帝国初代皇帝による人口土地調査

班田収授法

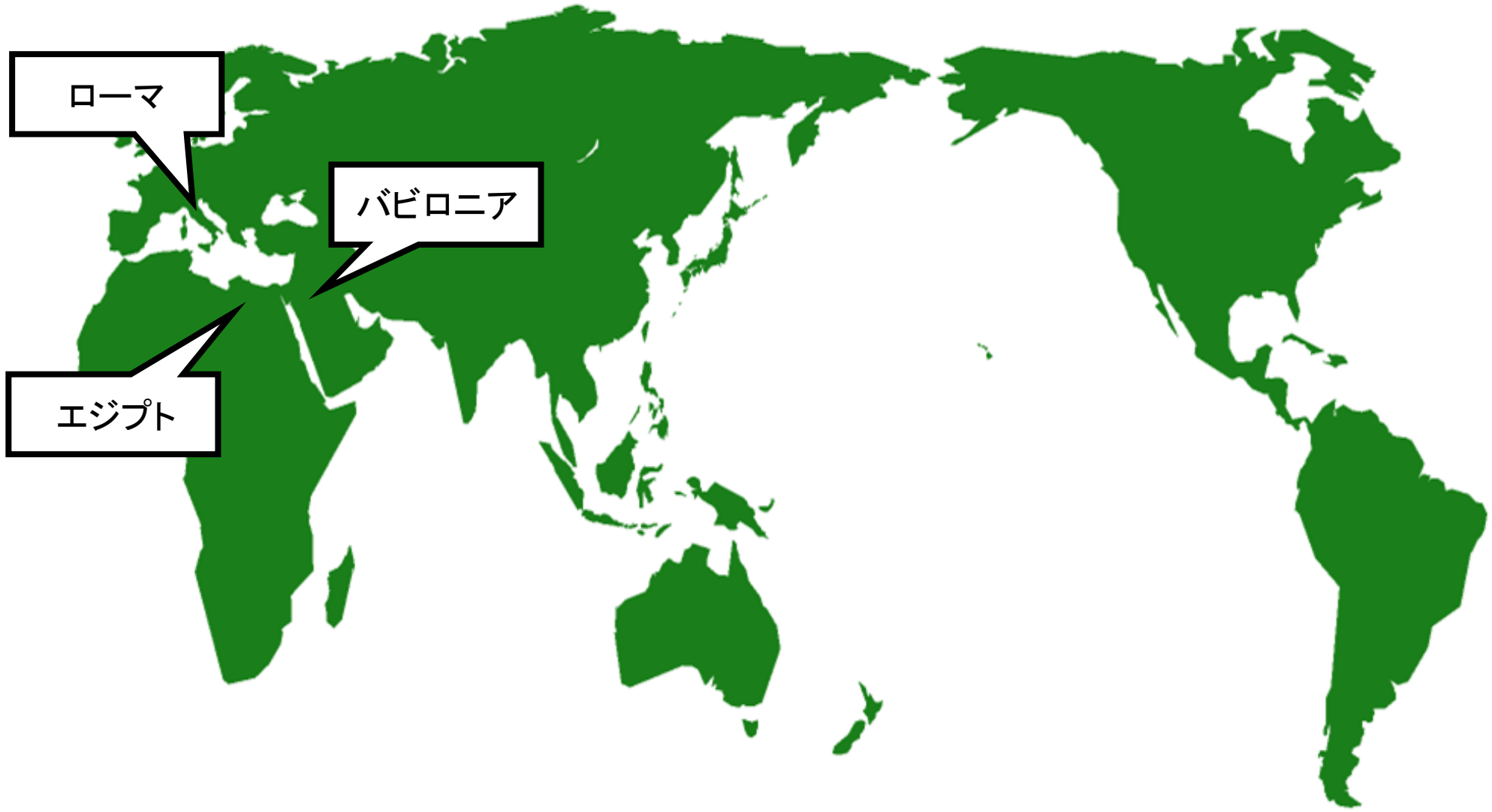


ドイツでは「国勢学」として
調査・研究が行われるようになる

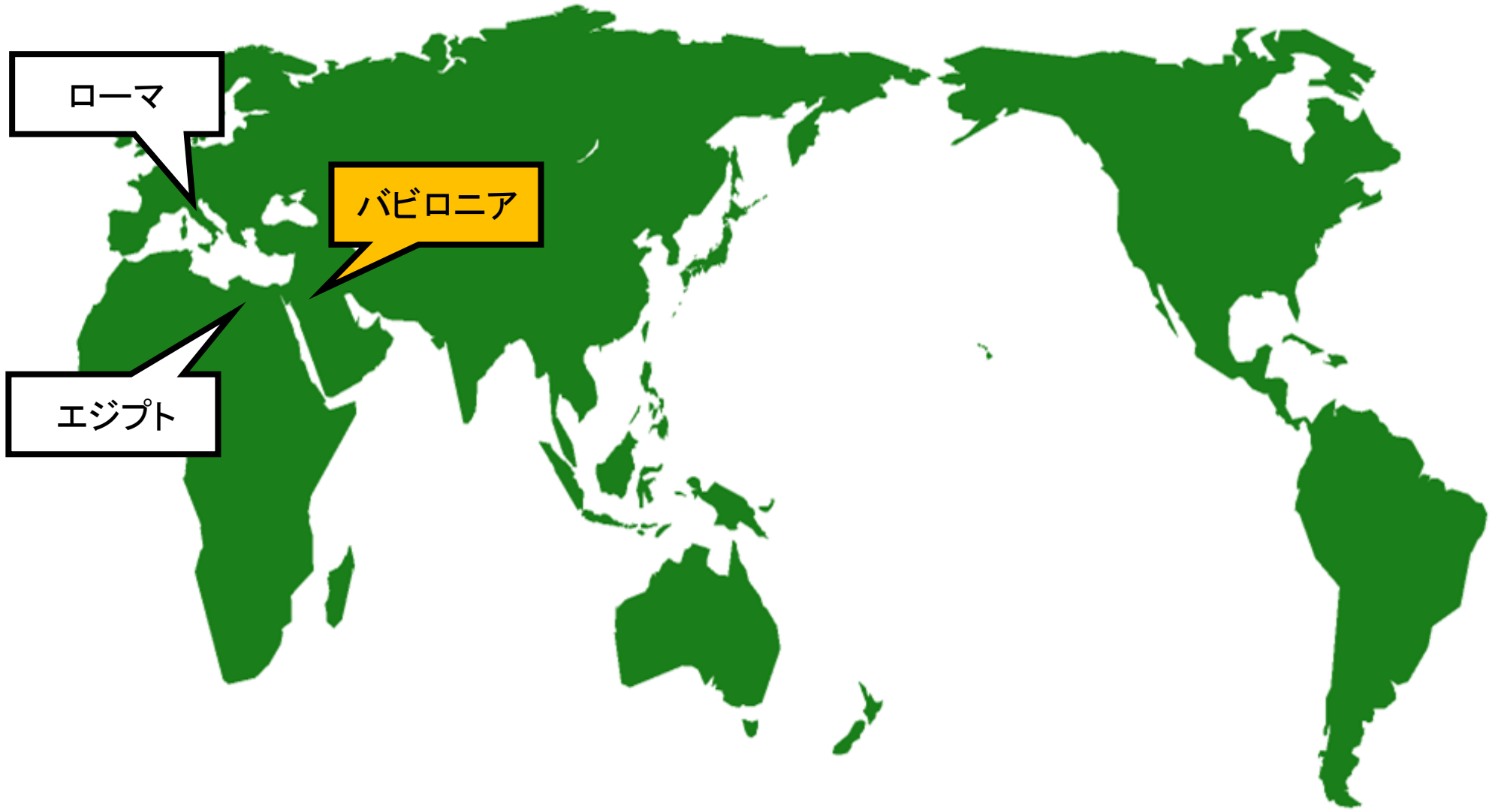
ナポレオンによって統計局が設置され
政府によって統計が整備される



集計の歴史



集計の歴史



集計の歴史

BC 4000年　バビロニア（メソポタミア文明）

集計の歴史

BC 4000年　バビロニア（メソポタミア文明）





集計の歴史

BC 4000年　バビロニア（メソポタミア文明）

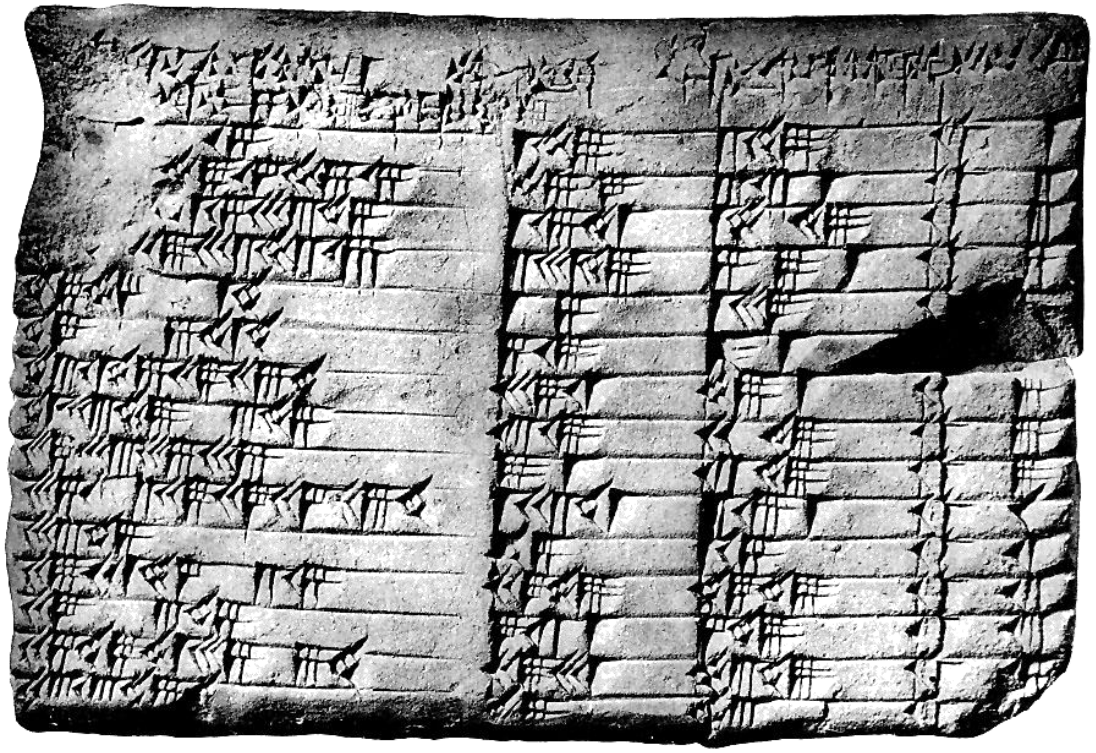




食糧と人間が釣り合っているか？

集計の歴史

BC 4000年　バビロニア（メソポタミア文明）



「プリンストン322」

集計の歴史

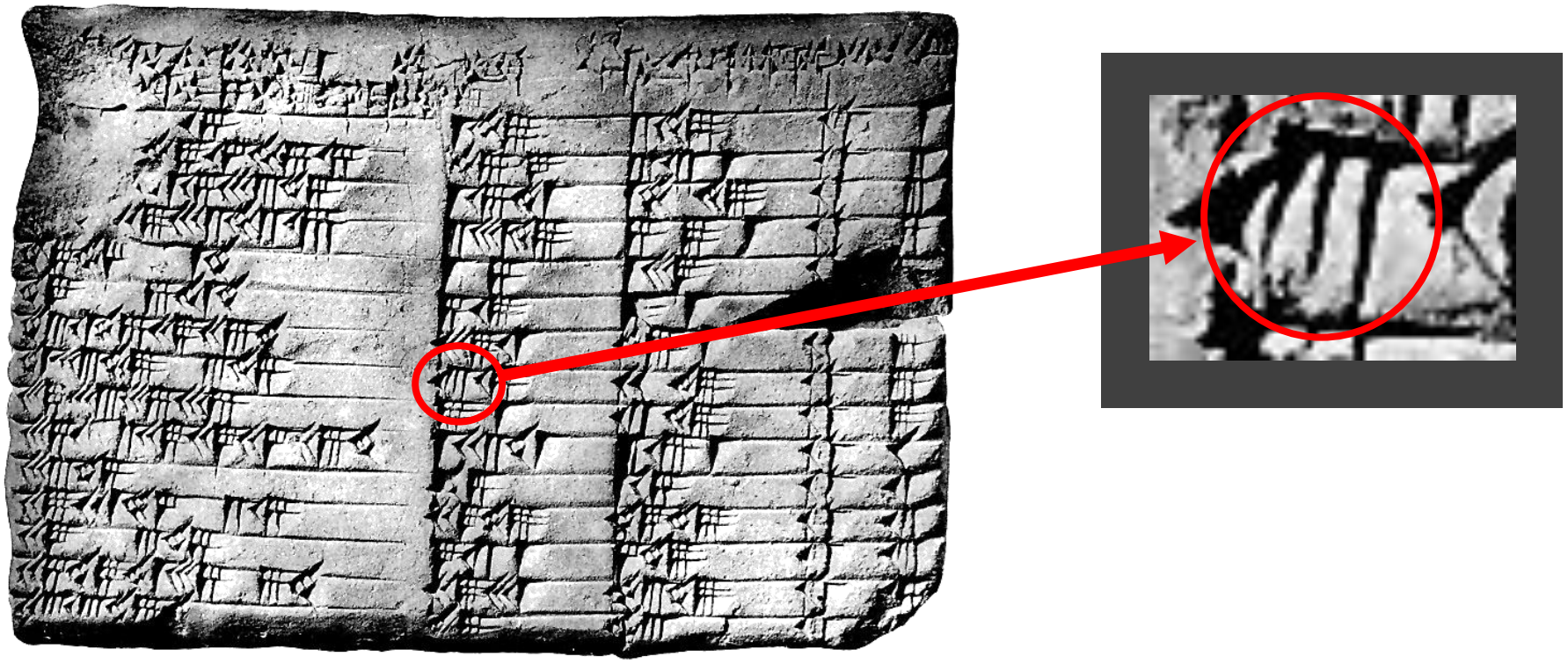
BC 4000年　バビロニア（メソポタミア文明）



「プリンストン322」

集計の歴史

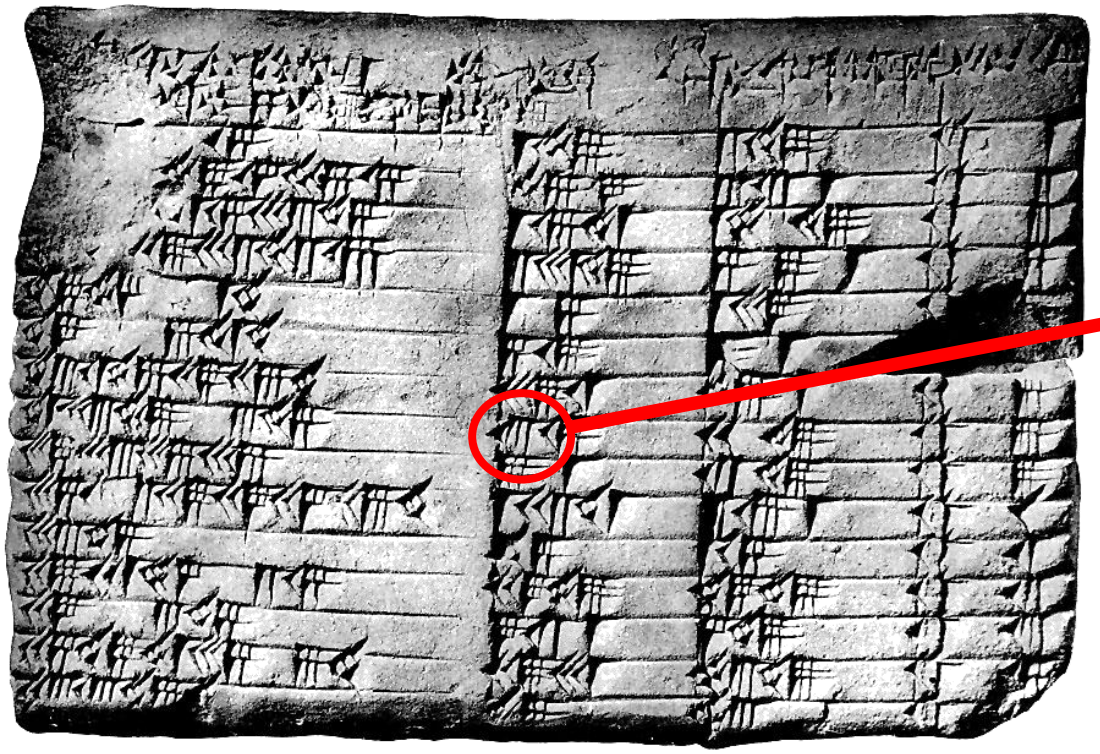
BC 4000年　バビロニア（メソポタミア文明）

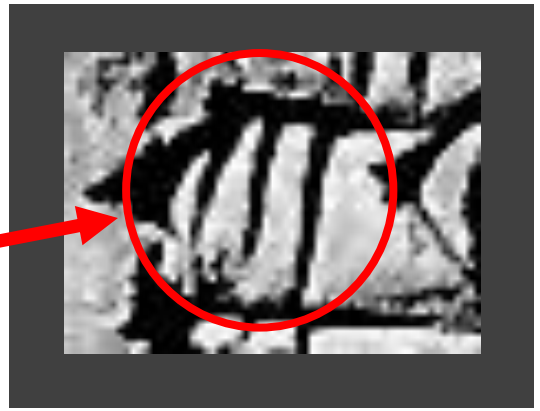


「プリンストン322」

集計の歴史

BC 4000年　バビロニア（メソポタミア文明）



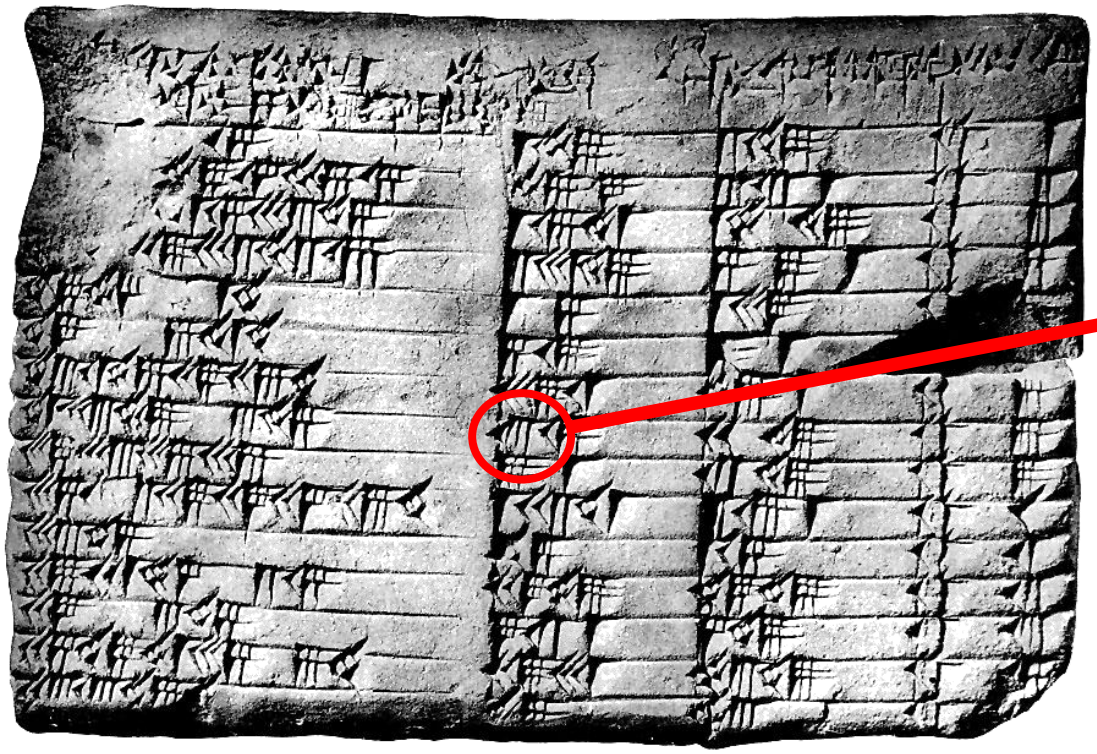




「プリンストン322」

集計の歴史

BC 4000年　バビロニア（メソポタミア文明）



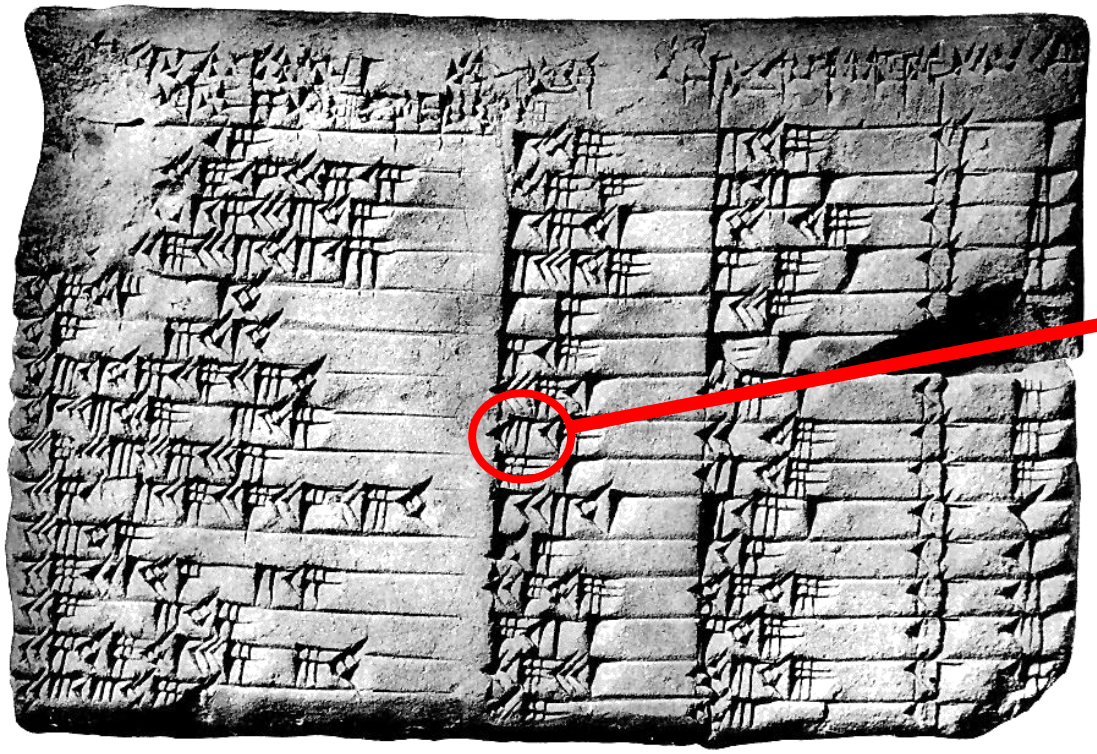




「プリンストン322」

集計の歴史

BC 4000年　バビロニア（メソポタミア文明）



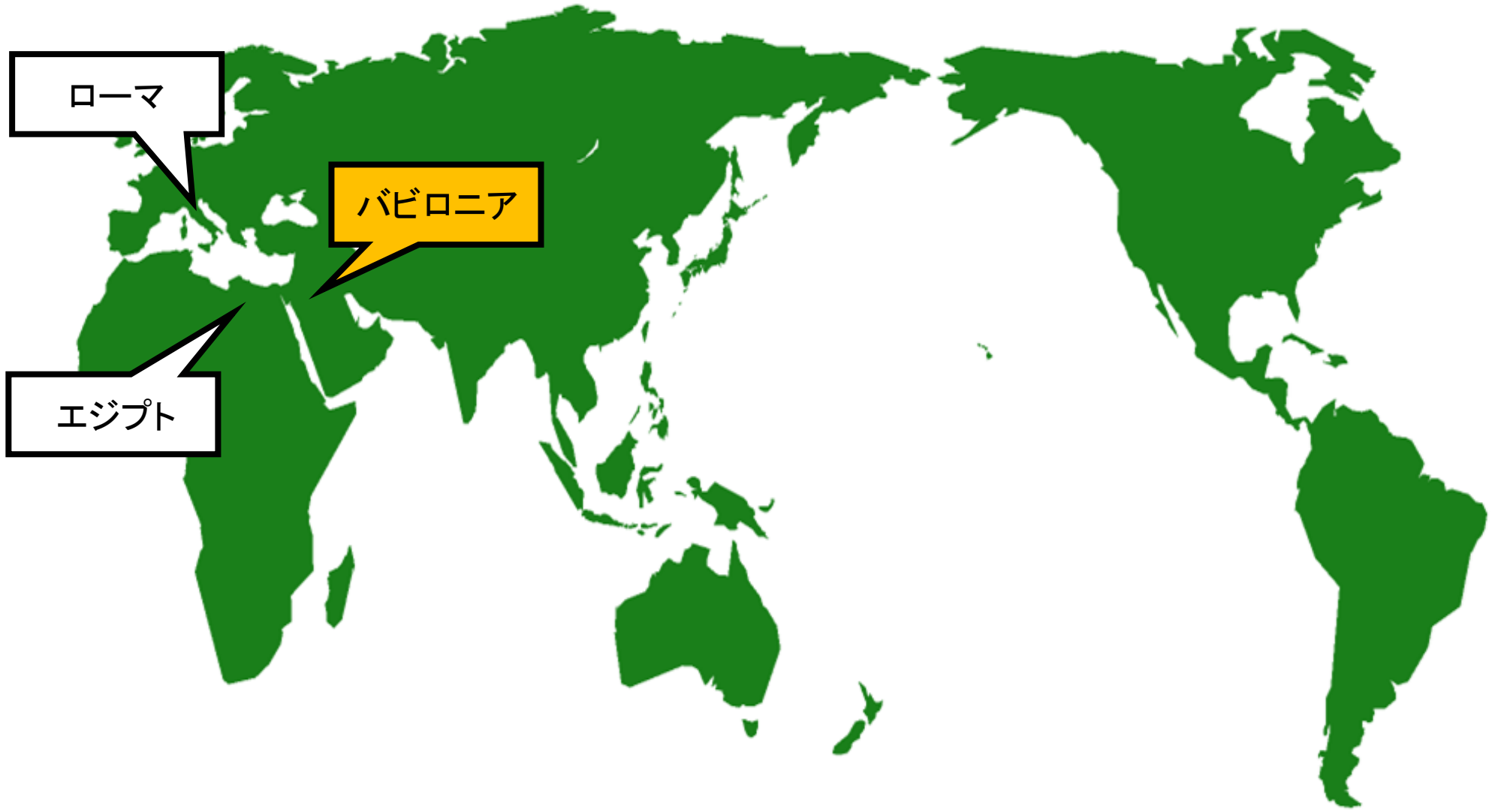
「プリンストン322」



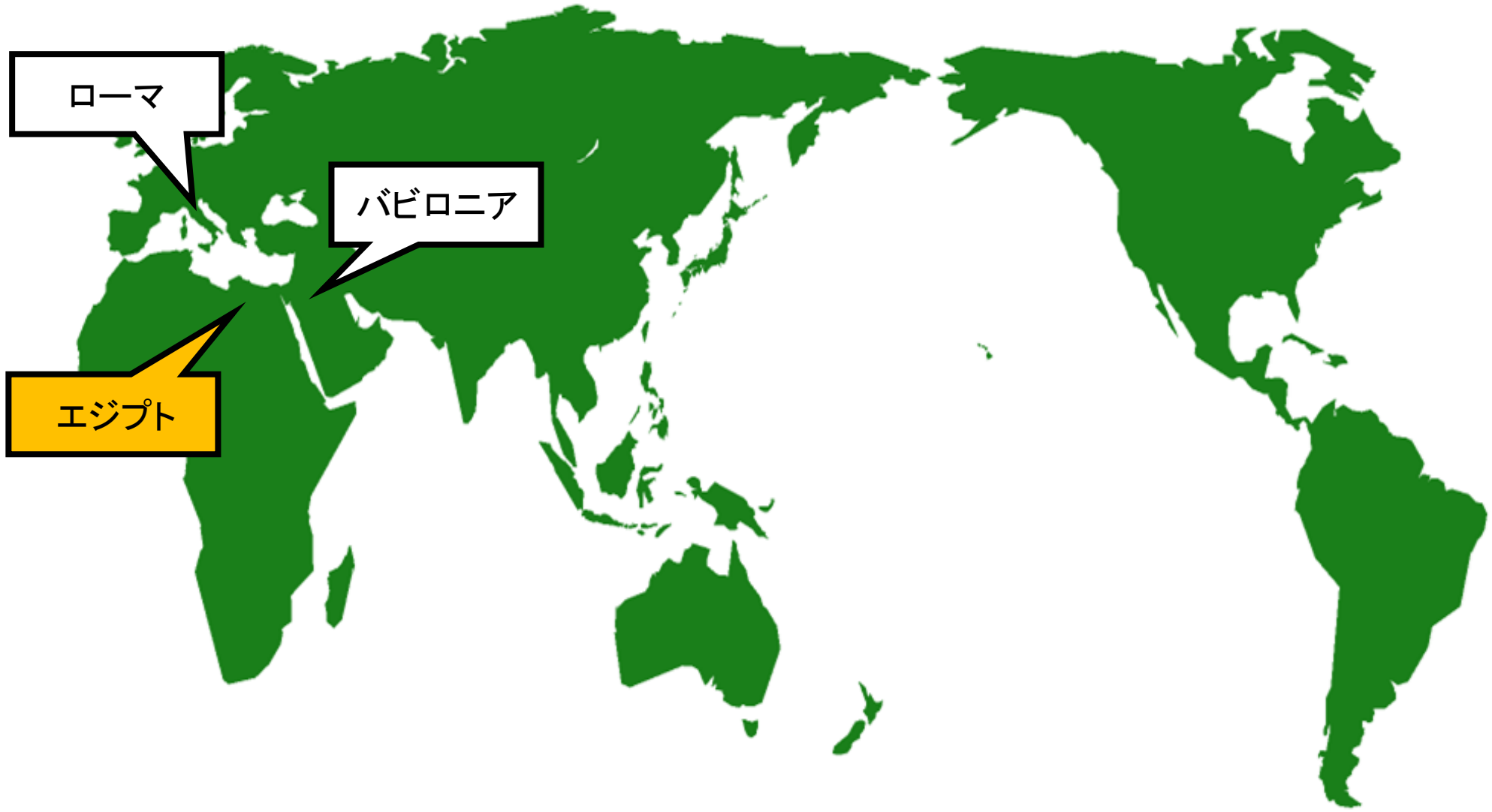
“3”



集計の歴史



集計の歴史



集計の歴史

BC 2500年　古代エジプト

ピラミッドの建設



集計の歴史

BC 2500年　古代エジプト

ピラミッドの建設



資材はどの程度必要か？



集計の歴史

BC 2500年　古代エジプト

ピラミッドの建設



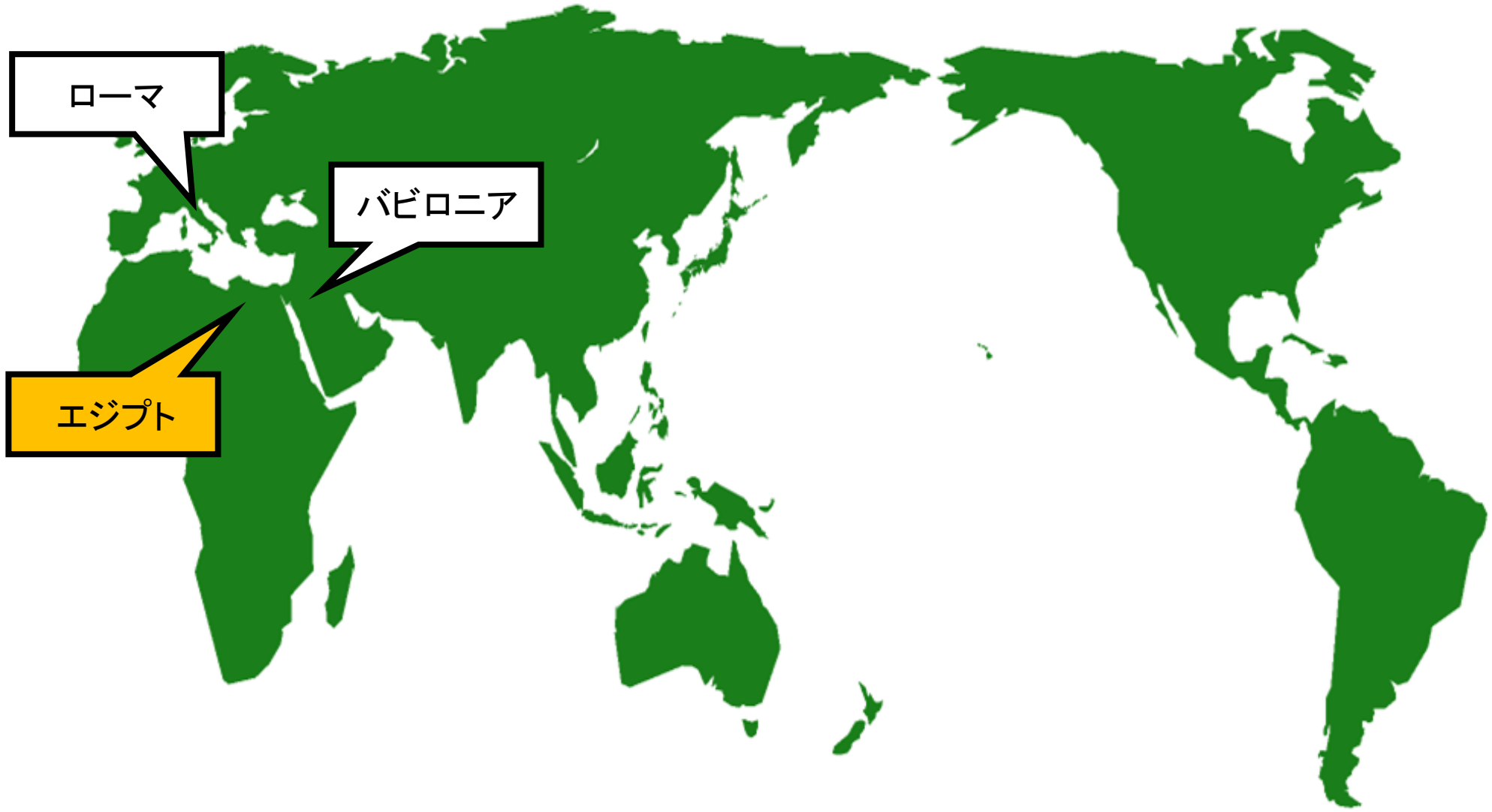
資材はどの程度必要か？



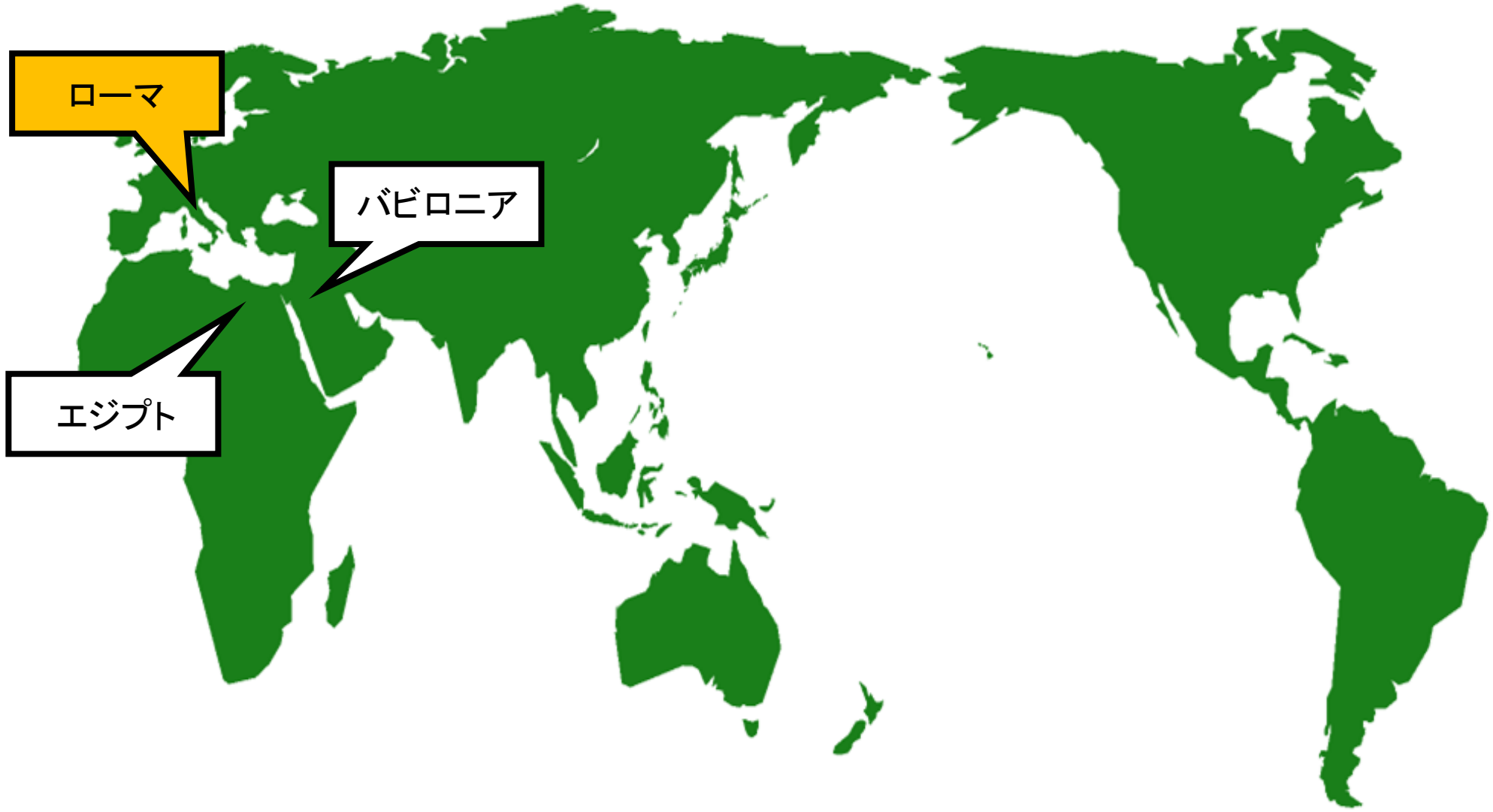
労働力はどの程度必要か？



集計の歴史



集計の歴史



集計の歴史

BC 200年　古代ローマ



集計の歴史

BC 200年　古代ローマ



兵隊は何人用意できるか？



集計の歴史

BC 200年　古代ローマ



兵隊は何人用意できるか？



税金をいくらまで取れるか？



現代統計学の大きな3つの源流

1. 国の実態を捉えるための「統計」
2. 確率的事象を捉えるための「統計」
3. 大量の事象を捉えるための「統計」



現代統計学の大きな3つの源流

1. 国の実態を捉えるための「統計」
2. 確率的事象を捉えるための「統計」
3. 大量の事象を捉えるための「統計」



2. 確率的事象を捉えるための「統計」

1500年　1600年　1700年　1800年

中世　近世

ジロラモ・カルダーノ
サイコロ賭博やトランプゲームにおける
偶然の仕組みの研究

ガリレオガリレイ
サイコロゲームについての考察」

パスカル、フェルマー
サイコロ賭博の問題について考察し
確率論の基礎を構築
（期待値・推定・検定・標本の理論など）

ベイズ理論

ド・モアブルの年金論

ラプラスの確率論



ラプラス

2. 確率的事象を捉えるための「統計」

1500年　1600年　1700年　1800年

中世　近世

ジロラモ・カルダーノ
サイコロ賭博やトランプゲームにおける
偶然の仕組みの研究

ガリレオガリレイ
サイコロゲームについての考察」

パスカル、フェルマー
サイコロ賭博の問題について考察し
確率論の基礎を構築
（期待値・推定・検定・標本の理論など）

ベイズ理論

ド・モアブルの年金論

ラプラスの確率論



ラプラス

ナポレオンの人口推計に応用

現代統計学の大きな3つの源流

1. 国の実態を捉えるための「統計」
2. 確率的事象を捉えるための「統計」
3. 大量の事象を捉えるための「統計」



３．大量の事象を捉えるための「統計」

1600年　　1700年　　1800年

近世

ペストに見舞われていたロンドンでジョン・グラントが死亡統計表を分析。

エドモンド・ハレーは死亡に一定の規律性があることを発見



エドモンド・ハレー

3. 大量の事象を捉えるための「統計」

1600年　　1700年　　1800年

近世

ペストに見舞われていたロンドンでジョン・グラントが死亡統計表を分析。

エドモンド・ハレーは死亡に一定の規律性があることを発見



エドモンド・ハレー

「偶然と見られる現象」に規律を
探求する手法としての統計

3. 大量の事象を捉えるための「統計」

1600年　1700年　1800年

近世

ペストに見舞われていたロンドンでジョン・グラントが死亡統計表を分析。

エドモンド・ハレーは死亡に一定の規律性があることを発見

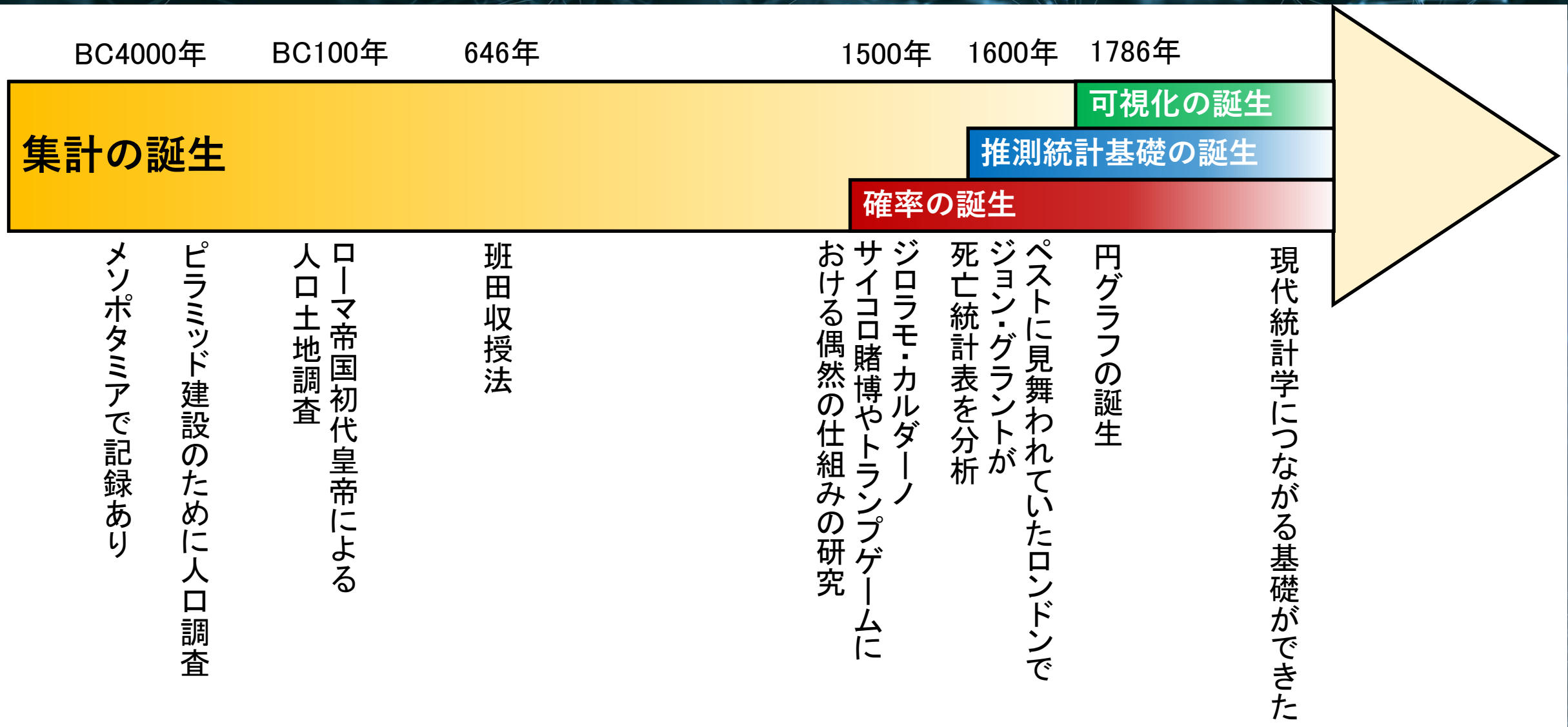


エドモンド・ハレー

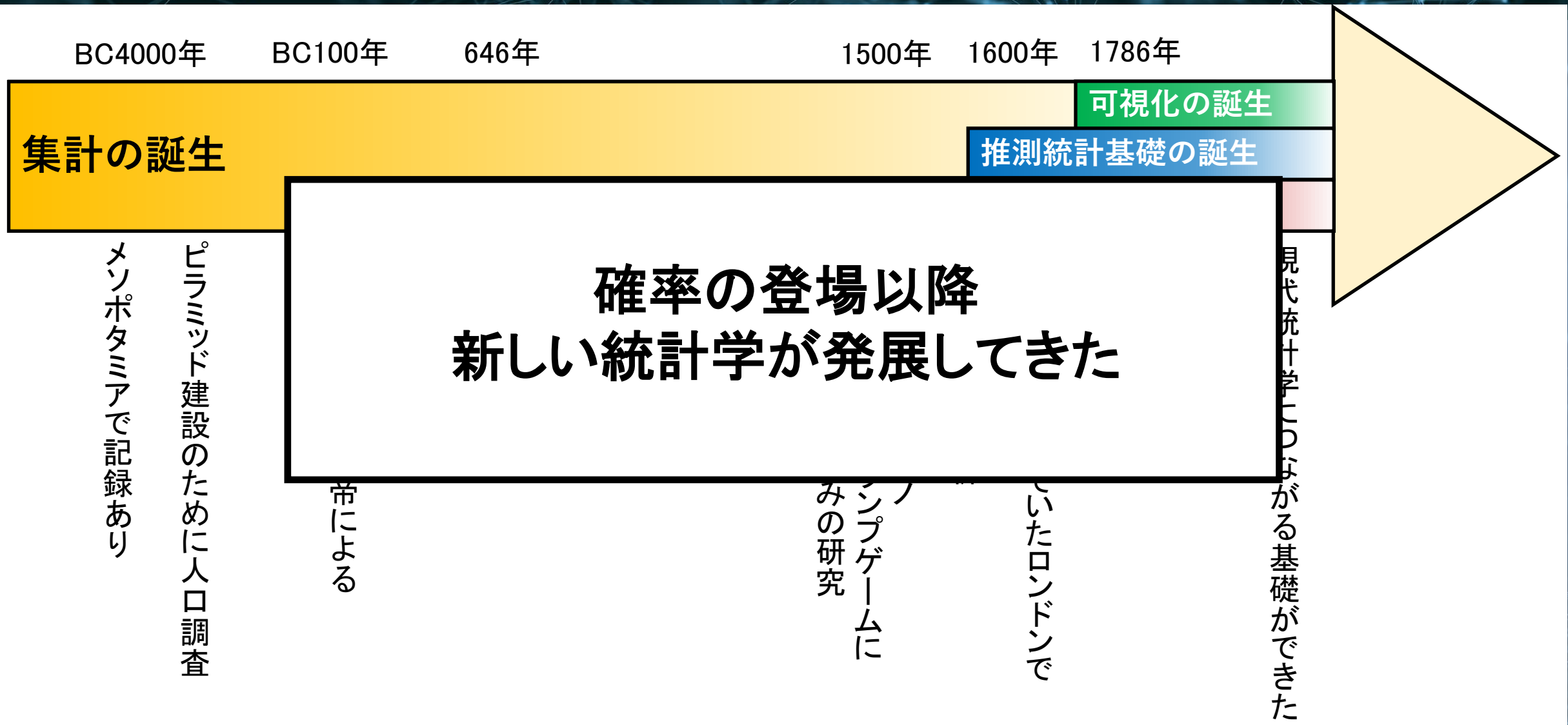
「偶然と見られる現象」に規律を探求する手法としての統計

“母集団”や“標本”の概念

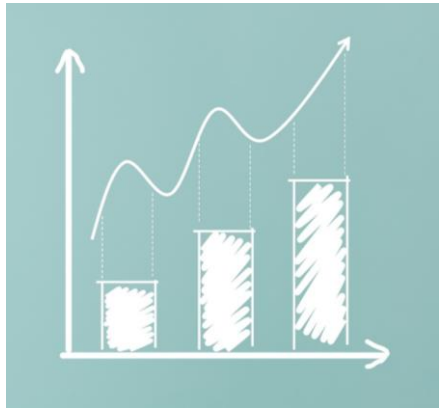
集計の歴史


BC4000年
BC100年
646年
1500年
1600年
1786年
集計の誕生
可視化の誕生
推測統計基礎の誕生
確率の誕生
メソポタミアで記録あり
ピラミッド建設のために人口調査
ローマ帝国初代皇帝による人口土地調査
班田収授法
ジロラモ・カルダーノサイコロ賭博やトランプゲームにおける偶然の仕組みの研究
ペストに見舞われていたロンドンでジョン・グラントが死亡統計表を分析
円グラフの誕生
現代統計学につながる基礎ができた

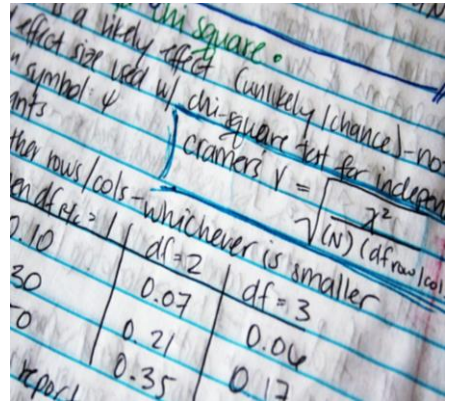
集計の歴史


BC4000年
BC100年
646年
1500年
1600年
1786年
集計の誕生
可視化の誕生
推測統計基礎の誕生
確率の登場以降
新しい統計学が発展してきた
メソポタミアで記録あり
ピラミッド建設のために人口調査
帝による
ンプゲームに
みの研究
いたロンドンで
現代統計学につながる基礎ができた

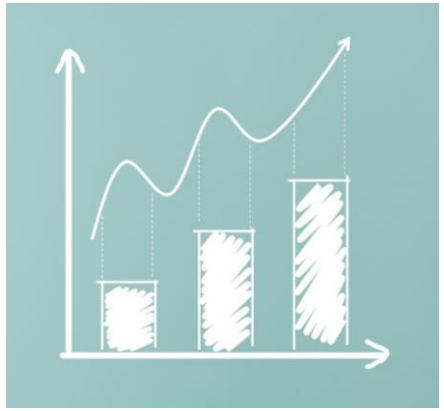
統計学の分類



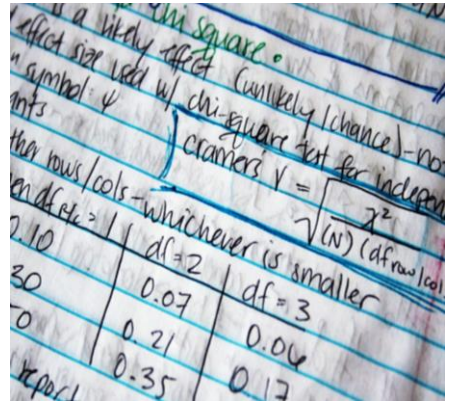
統計学


chi square
cramers
whichever is smaller
df=2
df=3
0.07
0.21
0.35
0.04

統計学の分類



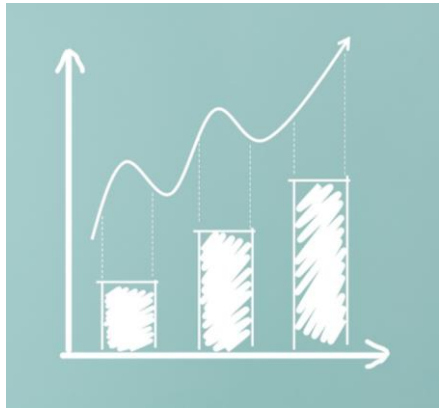
統計学



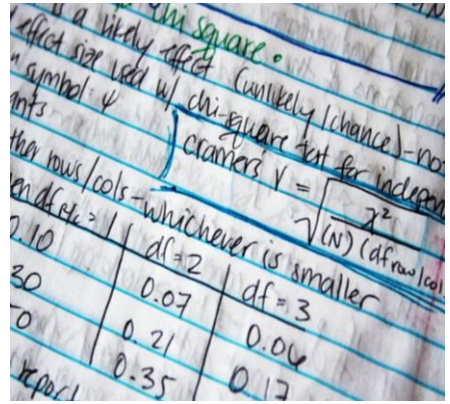
記述統計学

推測統計学

統計学の分類



統計学



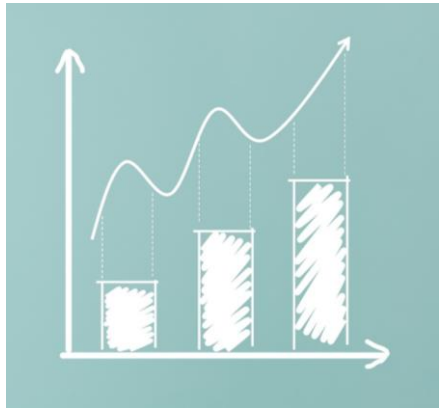
記述統計学

データを整理し、数値や表、
あるいはグラフなどを用いて
データの特徴を捉える

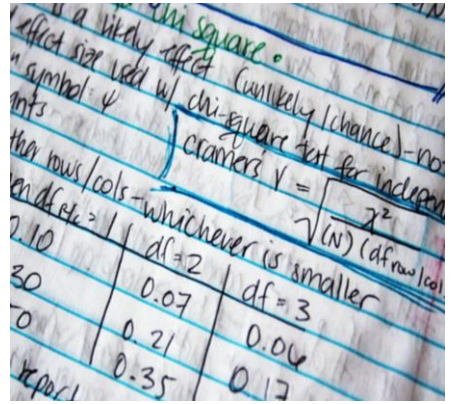
推測統計学

サンプルデータ（標本）から全体（母集団）の状況を推測する
推定 ・検定 ・ 相関・回帰分析

統計学の分類



統計学



記述統計学

データを整理し、数値や表、
あるいはグラフなどを用いて
データの特徴を捉える

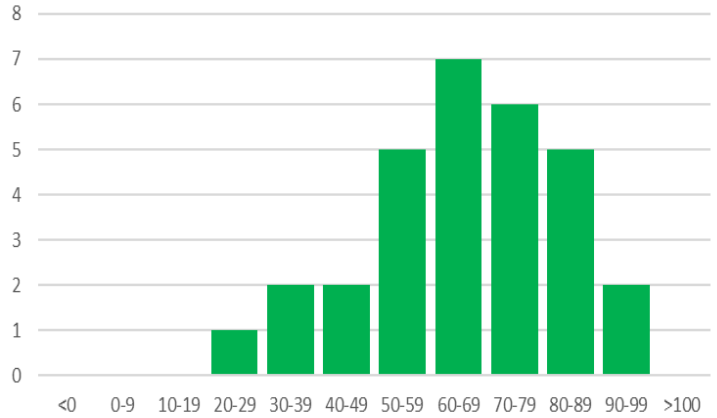
推測統計学

サンプルデータ（標本）から全体（母集団）の状況を推測する
推定 ・検定 ・ 相関・回帰分析

記述統計

問　A高校3年1組30人の数学の得点のデータを調べてください。

出席番号	得点
1	88
2	78
3	59
4	90
5	71
⋮	⋮
30	45



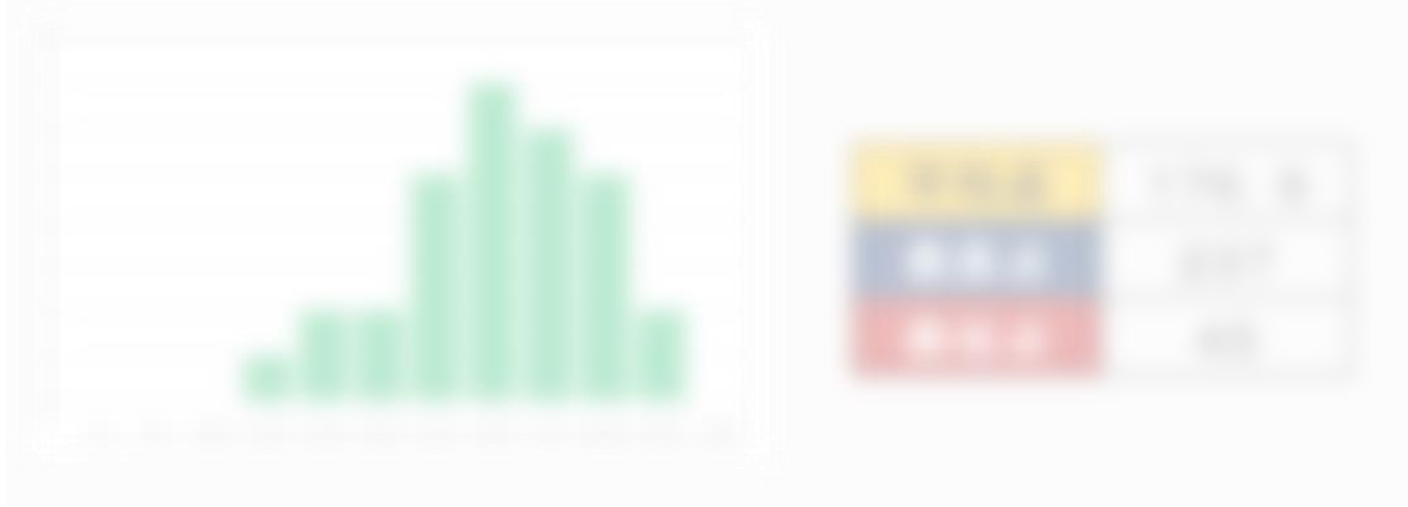
平均点	65．9
最高点	94
最低点	23

30人分のデータを取って特徴を調べる

記述統計

問　日本の1か月、1世帯あたりの牛肉の消費量を調べてください。

No.	消費量
1	1600
2	1550
3	1740
4	1810
5	1630
⋮	⋮
5000万	1590



日本国全世帯のデータを取って調べる…？

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



① データを取得

標本データ



記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



平均 μ

① データを取得

標本データ



② 標本データの傾向を調べる

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



平均 μ

① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

消費量平均	1605 円

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



平均 μ

① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

消費量平均	1605 円

③ 全体を推測する

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



平均 μ

① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

消費量平均	1605 円

③ 全体を推測する

→ **推測統計**

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



平均 μ

① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

消費量平均	1605 円

一部のデータから全体の傾向の推測が可能

③ 全体を推測する

→ **推測統計**

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

消費量平均	1605 円

一部のデータから全体の傾向の推測が可能

③ 全体を推測する

→ **推測統計**

例：家計調査、TVの視聴率調査、アンケート調査など

記述統計と推測統計

牛肉消費量を推定する

全世帯（母集団）



平均 μ

① データを取得

標本データ



② 標本データの傾向を調べる

→ **記述統計**

一部のデータから全体の傾向の推測が可能

③ 全体を推測する

→ **推測統計**

消費量平均	1605 円

例：家計調査、TVの視聴率調査、アンケート調査など

データのまとめ方（記述統計）

データのまとめ方

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128
2021/7/6	58
2021/7/7	91
2021/7/8	45
2021/7/9	126
2021/7/10	159

第１章データ.xlsx

どんな特徴がありますか？

これだけだと分かりづらい

データのまとめ方

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128
2021/7/6	58
2021/7/7	91
2021/7/8	45
2021/7/9	126
2021/7/10	159

第１章データ.xlsx

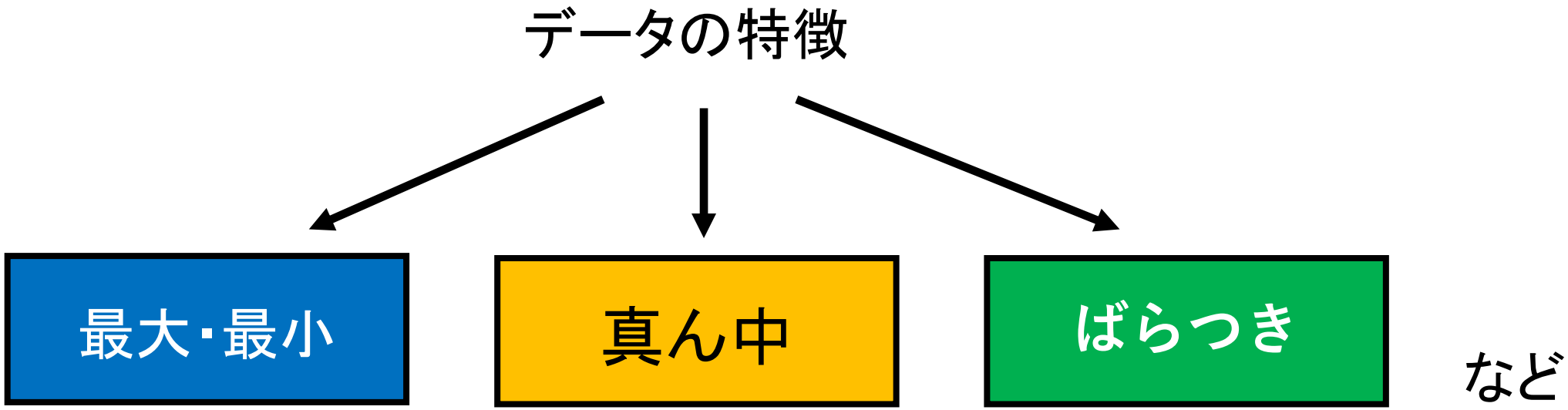
どんな特徴がありますか？

これだけだと分かりづらい



「統計量」を使ってデータの特徴を見てみましょう。

データの特徴をとらえる



データの特徴をとらえる（最大値・最小値）

最大値・最小値

データの中で一番大きな/小さな値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

データのとりうる範囲がわかる

統計量	値
最大値	341
最小値	3

データの特徴をとらえる（最大値・最小値）

最大値・最小値

データのとりうる範囲がわかる

データの中で一番大きな/小さな値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
最大値	9999
最小値	-100

最大値・最小値を計算する意味

最小値、最大値に「ありえない値」が出てきた場合、
データにエラーがある可能性があります。

データの特徴をとらえる（平均値）

平均値

データを平らに均した値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

データの概ね「真ん中」と解釈できる指標

統計量	値
最大値	341
最小値	3

データの特徴をとらえる（平均値）

平均値

データの概ね「真ん中」と解釈できる指標

データを平らに均した値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
最大値	341
最小値	3

$$(106 + 113 + 201 + 320 + 128 + \cdots + 99) \div 120$$

データの特徴をとらえる（平均値）

平均値

データの概ね「真ん中」と解釈できる指標

データを平らに均した値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
最大値	341
最小値	3

$$(106 + 113 + 201 + 320 + 128 + \cdots + 99) \div 120$$

約3か月間の合計販売量　　データ数

データの特徴をとらえる（平均値）

平均値

データの概ね「真ん中」と解釈できる指標

データを平らに均した値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
最小値	3

$$(106 + 113 + 201 + 320 + 128 + \cdots + 99) \div 120 = \mathbf{125.8}$$

約3か月間の合計販売量　　データ数

データの特徴をとらえる（中央値）

中央値（メディアン） 平均値同様に概ね「真ん中」と解釈できる指標

データを小さい（大きい）順に並べたときの中心の値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
中央値	109.5
最小値	3

データの特徴をとらえる（中央値）

中央値（メディアン） 平均値同様に概ね「真ん中」と解釈できる指標

データを小さい（大きい）順に並べたときの中心の値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
中央値	109.5
最小値	3

データの特徴をとらえる（中央値）

中央値（メディアン）

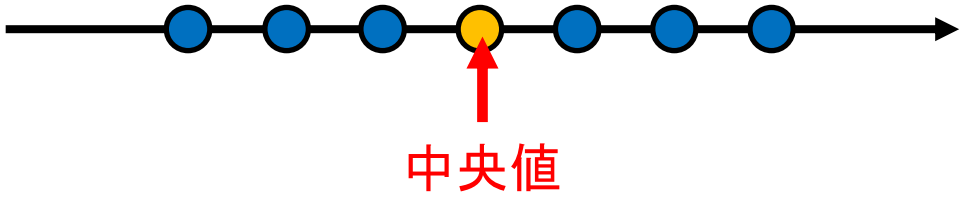
平均値同様に概ね「真ん中」と解釈できる指標

データを小さい（大きい）順に並べたときの中心の値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
中央値	109.5
最小値	3

データ数が奇数個の場合



データの特徴をとらえる（中央値）

中央値（メディアン）

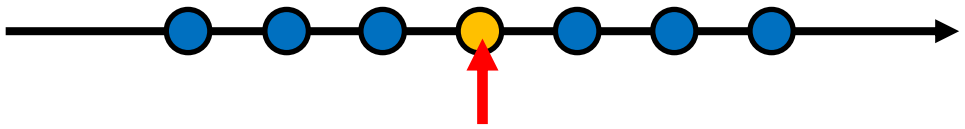
平均値同様に概ね「真ん中」と解釈できる指標

データを小さい（大きい）順に並べたときの中心の値

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

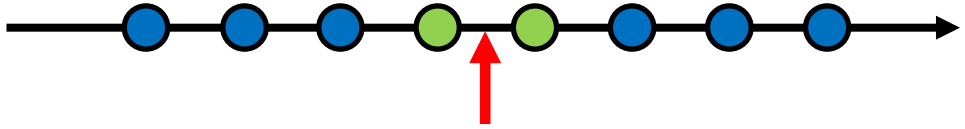
統計量	値
平均値	125.8
最大値	341
中央値	109.5
最小値	3

データ数が奇数個の場合



中央値

データ数が偶数個の場合



中央値

データの特徴をとらえる（四分位数）

25%点, 75%点

データを4分割する指標

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

75%点：中央値と最大値の中央
25%点：最小値と中央値の中央

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3

データの特徴をとらえる（四分位数）

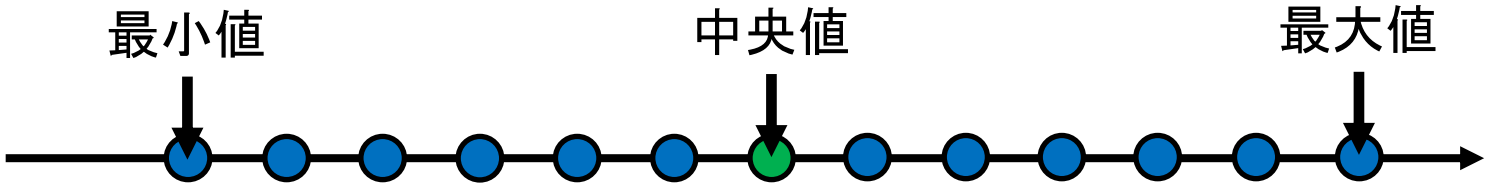
25%点，75%点

データを4分割する指標

75%点：中央値と最大値の中央
25%点：最小値と中央値の中央

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3



データの特徴をとらえる（四分位数）

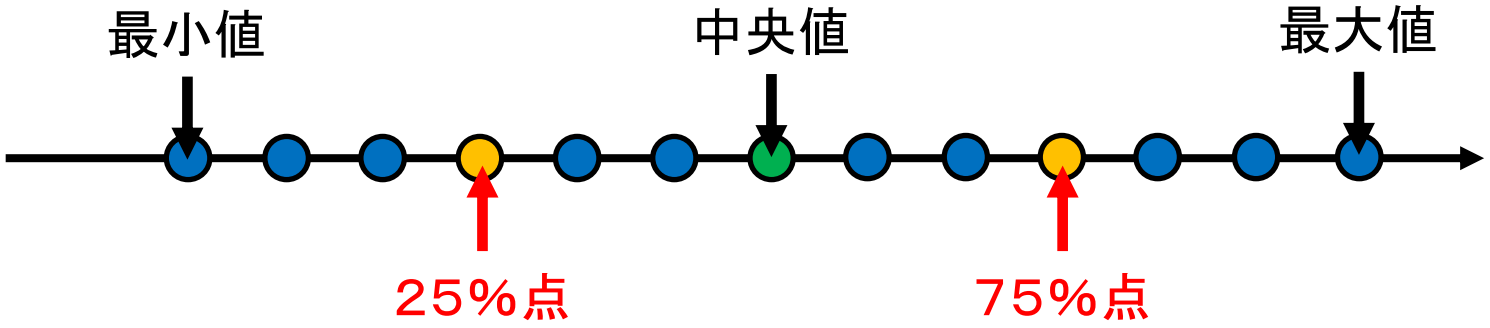
25%点，75%点

データを4分割する指標

75%点：中央値と最大値の中央
25%点：最小値と中央値の中央

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3



データの特徴をとらえる（標準偏差）

標準偏差

データの散らばり具合を数値化

平均からどれだけばらついているかを測る指標

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

データの特徴をとらえる（標準偏差）

標準偏差

データの散らばり具合を数値化

平均からどれだけばらついているかを測る指標

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

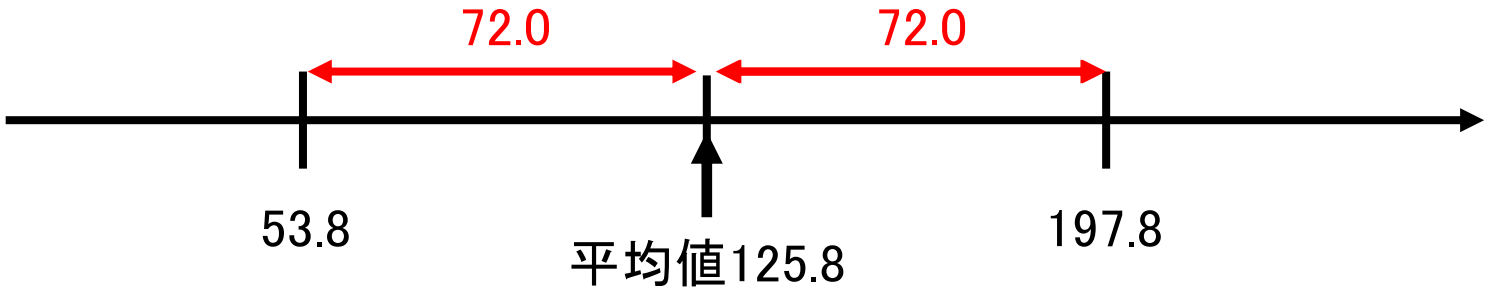
データの特徴をとらえる（標準偏差）

標準偏差

データの散らばり具合を数値化

平均からどれだけばらついているかを測る指標

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128



統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

データの特徴をとらえる（標準偏差）

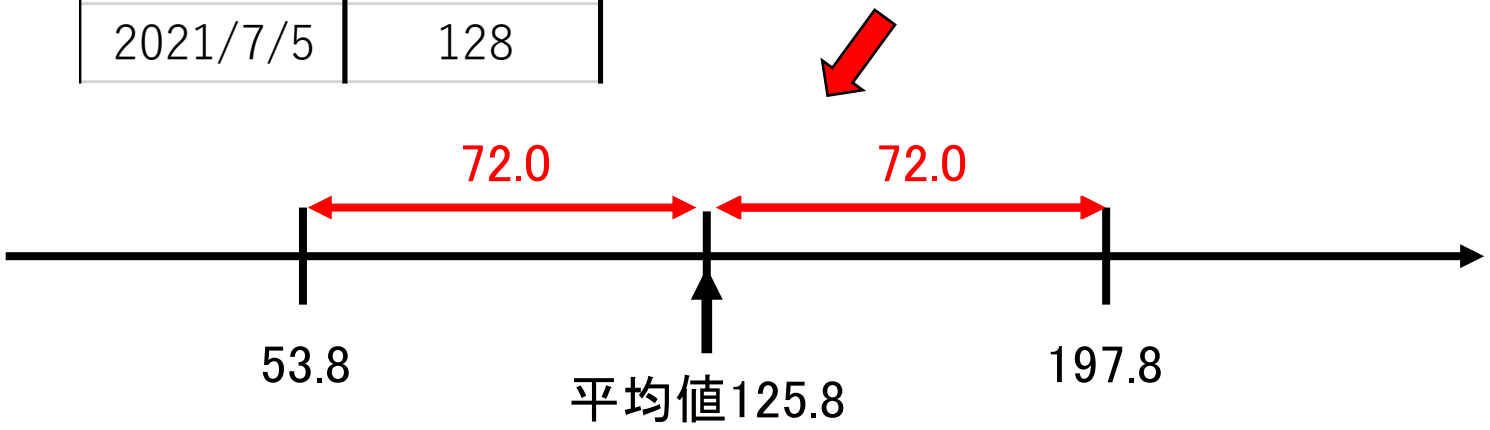
標準偏差

データの散らばり具合を数値化

平均からどれだけばらついているかを測る指標

日付	Aの販売量
2021/7/1	106
2021/7/2	113
2021/7/3	201
2021/7/4	320
2021/7/5	128

平均値から±72.0の範囲にデータの概ねが入っていると解釈できます。



統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

Excel関数

基本的な統計量をExcelで計算してみましょう。

統計量	Excel関数
平均値	=AVERAGE(データ)
最大値	=QUARTILE.INC(データ, 4) もしくは =MAX(データ)
75%点	=QUARTILE.INC(データ, 3)
中央値	=QUARTILE.INC(データ, 2) もしくは =MEDIAN(データ)
25%点	=QUARTILE.INC(データ, 1)
最小値	=QUARTILE.INC(データ, 0) もしくは =MIN(データ)
標準偏差	=STDEV.S(データ)

Excel実践

Excelハンズオン



各指標を読み解いてみよう

最大値・最小値を確認

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

各指標を読み解いてみよう

最大値・最小値を確認

統計量	値
平均値	125.8
最大値	**341**
７５％点	161.5
中央値	109.5
２５％点	82
最小値	**3**
標準偏差	72.0

データに異常・エラーはなさそう

各指標を読み解いてみよう

平均値を確認

統計量	値
平均値	**125.8**
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

各指標を読み解いてみよう

平均値を確認

統計量	値
平均値	**125.8**
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

→ 1日あたりだいたい126個ぐらい売れている

各指標を読み解いてみよう

標準偏差を確認

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	**72.0**

各指標を読み解いてみよう

標準偏差を確認

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	**72.0**

「ばらつき」が72とは？

各指標を読み解いてみよう

標準偏差を確認

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	**72.0**

「ばらつき」が72とは？

平均から±72の範囲に
概ねのデータが入っている

各指標を読み解いてみよう

標準偏差を確認

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	**72.0**

「ばらつき」が72とは？

⇩

平均から±72の範囲に
概ねのデータが入っている

⇩

126＋72＝198個在庫があれば
売り逃しはなさそう

各指標を読み解いてみよう

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

これでデータの特徴が把握できたといっていいでしょうか？

各指標を読み解いてみよう

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

これでデータの特徴が把握できたといっていいでしょうか？

NO！

各指標を読み解いてみよう

平均値の弱点

データA	データB
1	1
2	1
3	1
4	1
5	11

各指標を読み解いてみよう

平均値の弱点

データA	データB
1	1
2	1
3	1
4	1
5	11

平均値はともに3

各指標を読み解いてみよう

平均値の弱点

データA	データB
1	1
2	1
3	1
4	1
5	11

平均値はともに3

平均値が「真ん中」と解釈しにくいときもあります。

各指標を読み解いてみよう

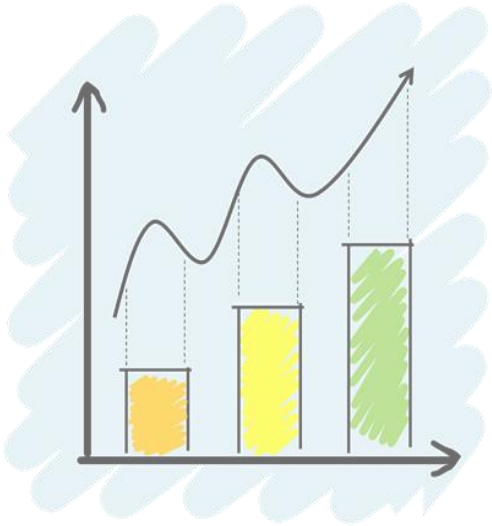
平均値の弱点

データA	データB
1	1
2	1
3	1
4	1
5	11

平均値はともに3

平均値が「真ん中」と解釈しにくいときもあります。

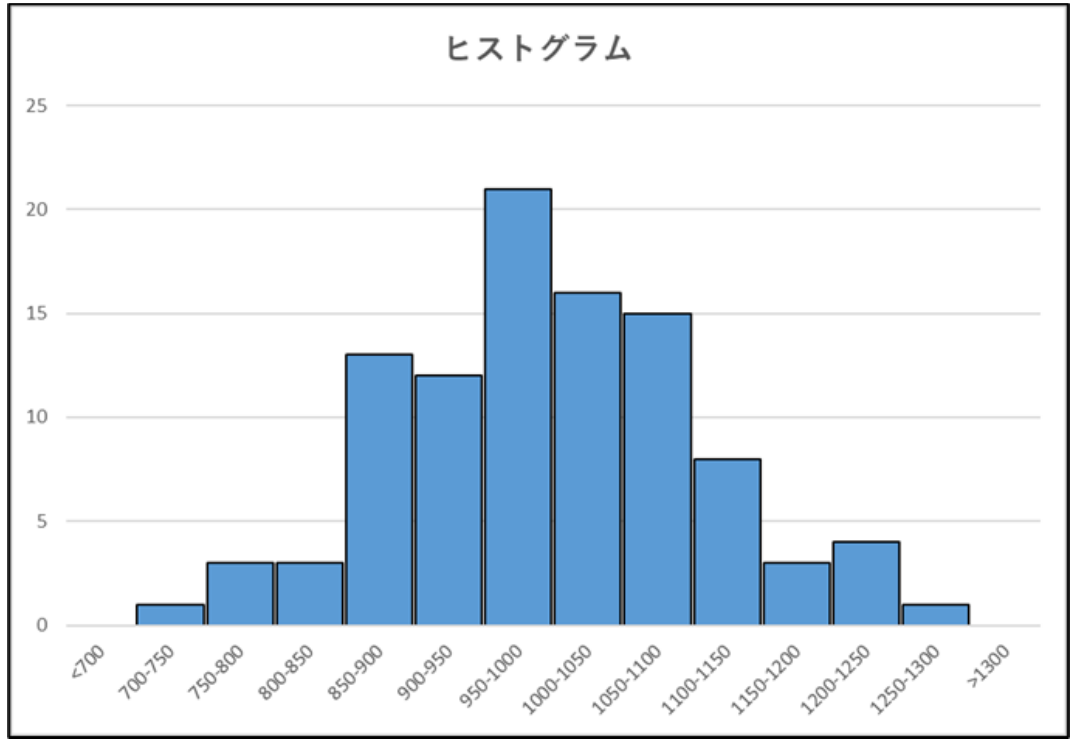
グラフを使ってデータを可視化する



ヒストグラム

ヒストグラム

データを区分(階級)分けし、それぞれの区分ごとにデータの個数を積み上げたグラフ



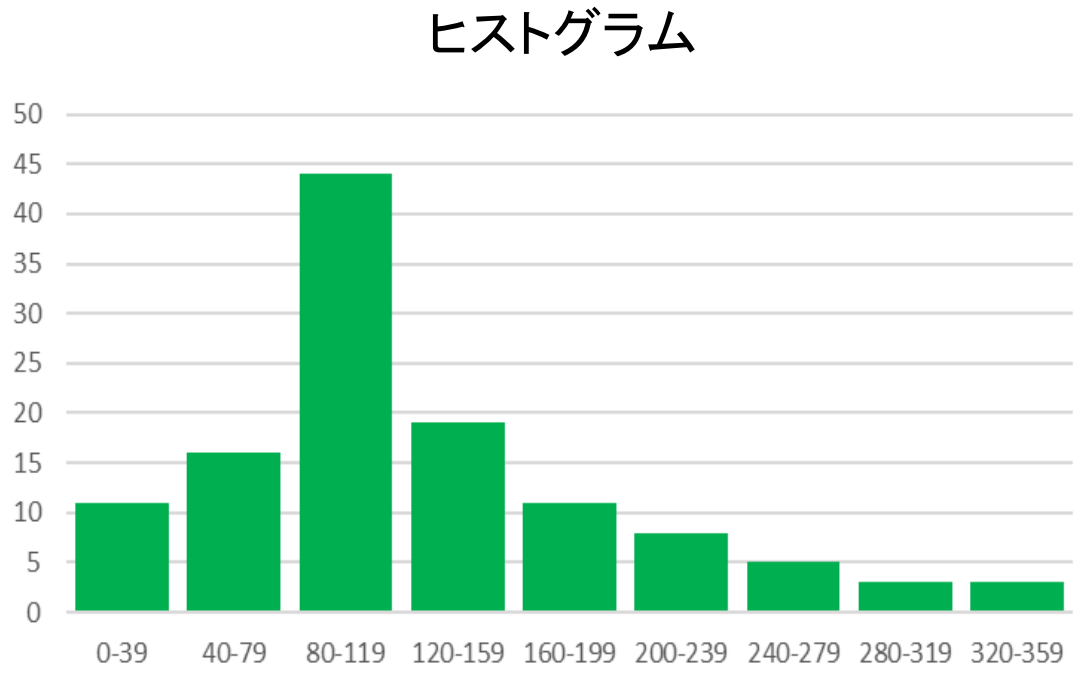
Excel実践

Excelでヒストグラムを描いてみよう



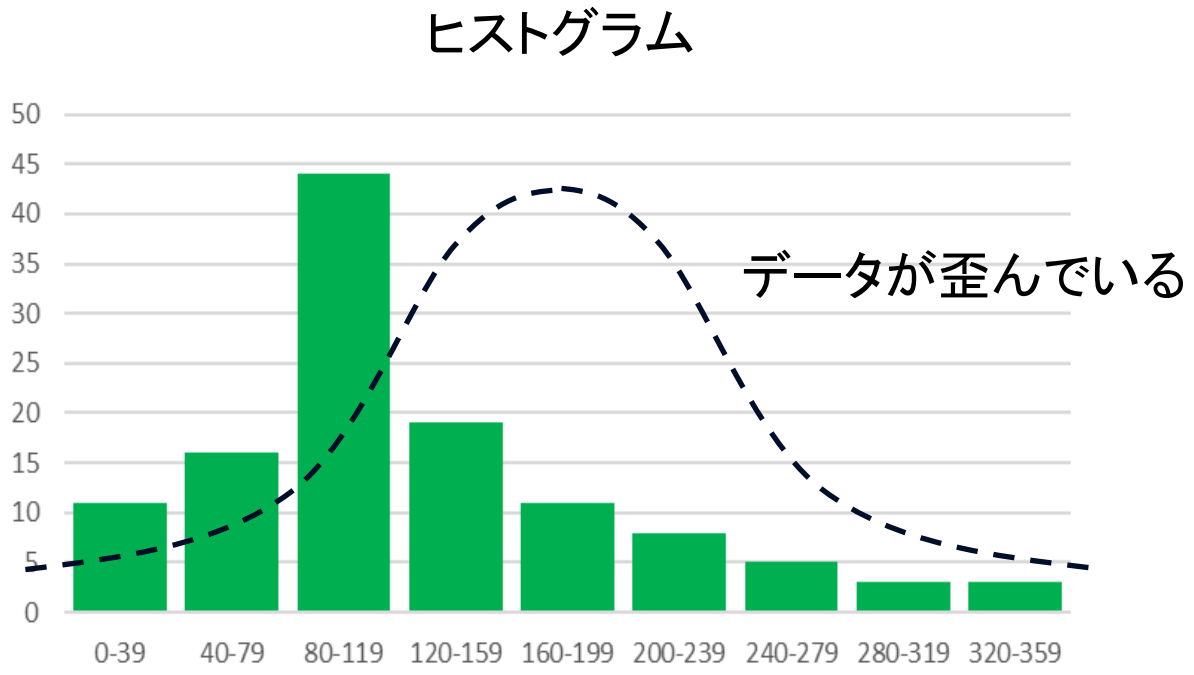
ヒストグラム

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0



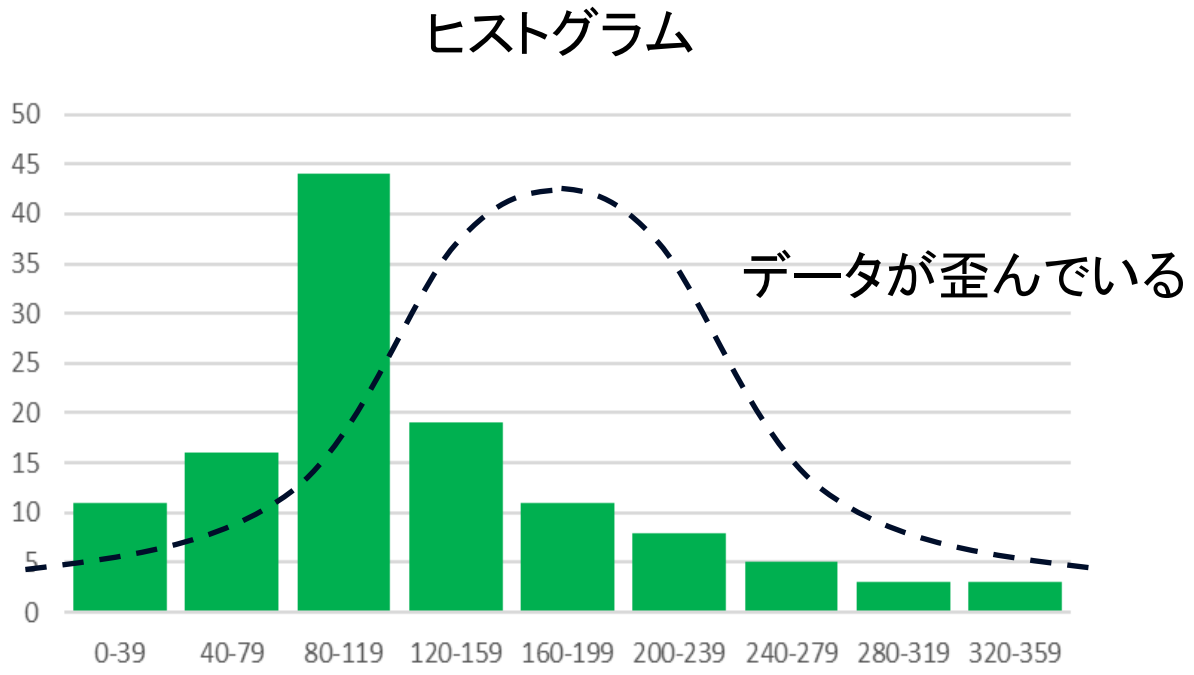
ヒストグラム

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0



ヒストグラム

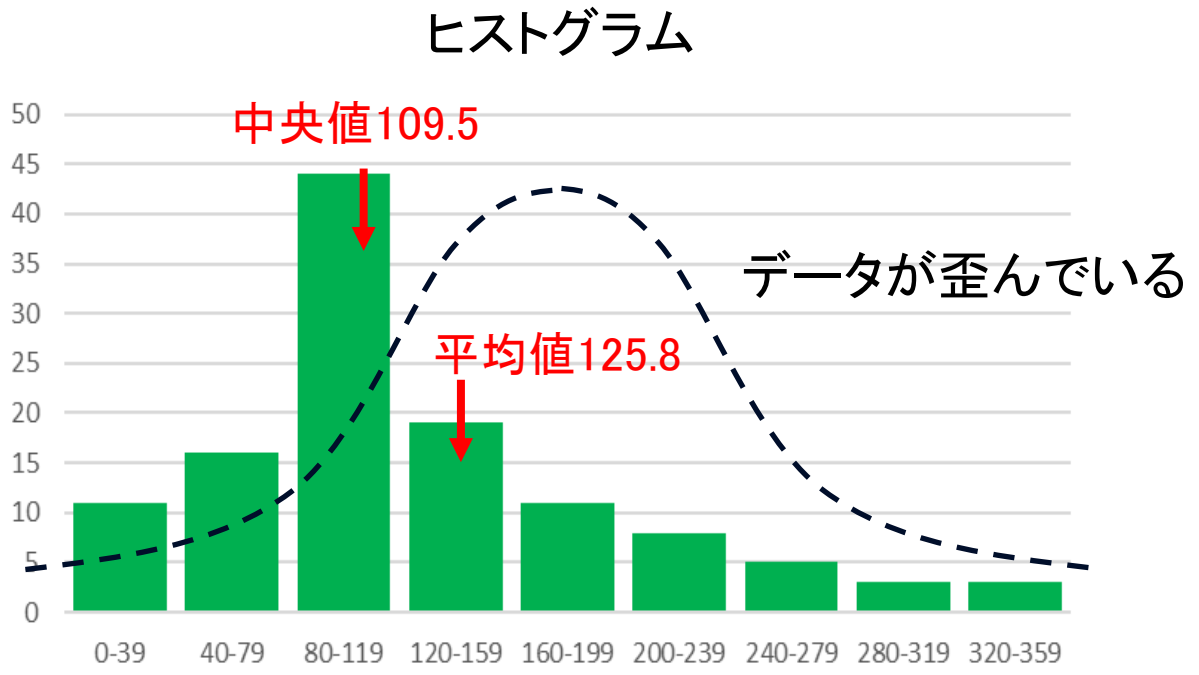
統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0



・大きな値がわずかにあるということ

ヒストグラム

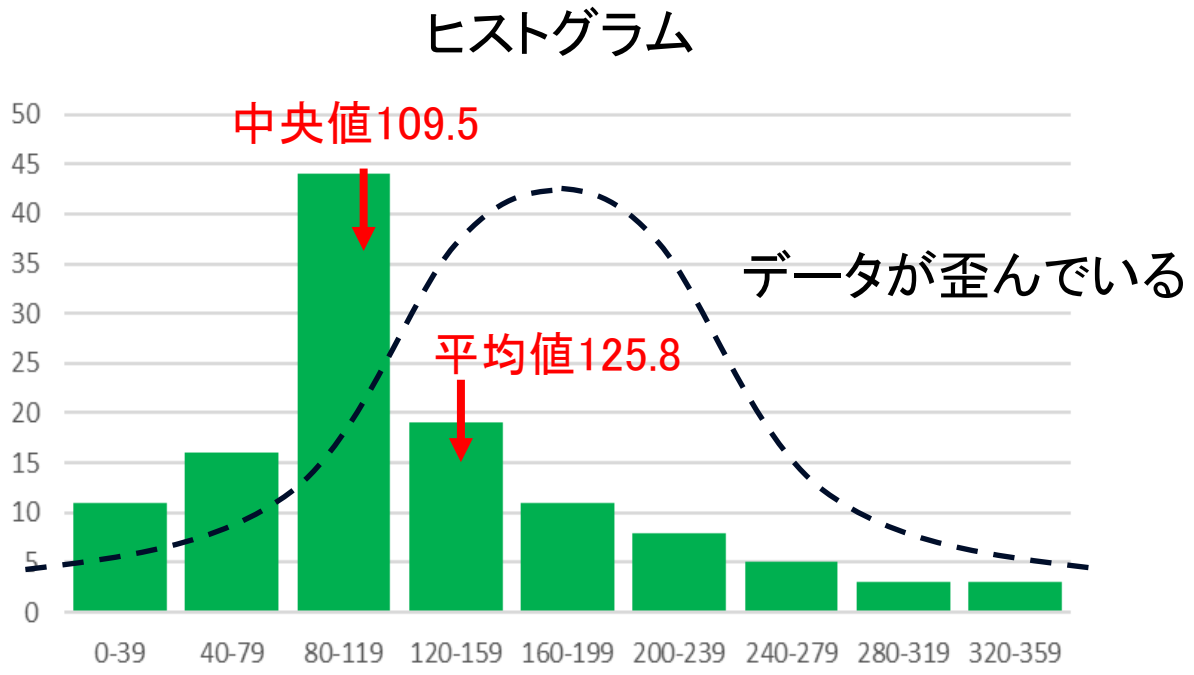
統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0



・大きな値がわずかにあるということ

ヒストグラム

統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0



- 大きな値がわずかにあるということ
- 平均値よりも中央値の方が「真ん中」と解釈しやすい

平均値がうまく機能しない場合

平均値がうまく機能しない場合

平均値がうまく機能しない場合

平均値がうまく機能しない場合

① データ数が極端に少ない場合

平均値がうまく機能しない場合

平均値がうまく機能しない場合

① データ数が極端に少ない場合

例：受験者数2名の平均点50点 → 0点と100点で平均50点

平均値がうまく機能しない場合

平均値がうまく機能しない場合

① データ数が極端に少ない場合

例：受験者数2名の平均点50点 → 0点と100点で平均50点

② 外れ値の影響を受ける場合

平均値がうまく機能しない場合

平均値がうまく機能しない場合

① データ数が極端に少ない場合

例：受験者数2名の平均点50点 → 0点と100点で平均50点

② 外れ値の影響を受ける場合

例：A社の平均年収600万円 → 極端に年収が高い社員が存在すれば平均値は引き上げられる

平均値がうまく機能しない場合

平均値がうまく機能しない場合

① データ数が極端に少ない場合

例：受験者数2名の平均点50点 → 0点と100点で平均50点

② 外れ値の影響を受ける場合

例：A社の平均年収600万円 → 極端に年収が高い社員が存在すれば平均値は引き上げられる

極端な値の影響を受けない「中央値」を見た方がよさそうです。

ヒストグラム

中央値の弱点

例：あるクラスの身長の中央値は160cm（平均値も160cm）

ヒストグラム

中央値の弱点

例：あるクラスの身長の中央値は160cm（平均値も160cm）

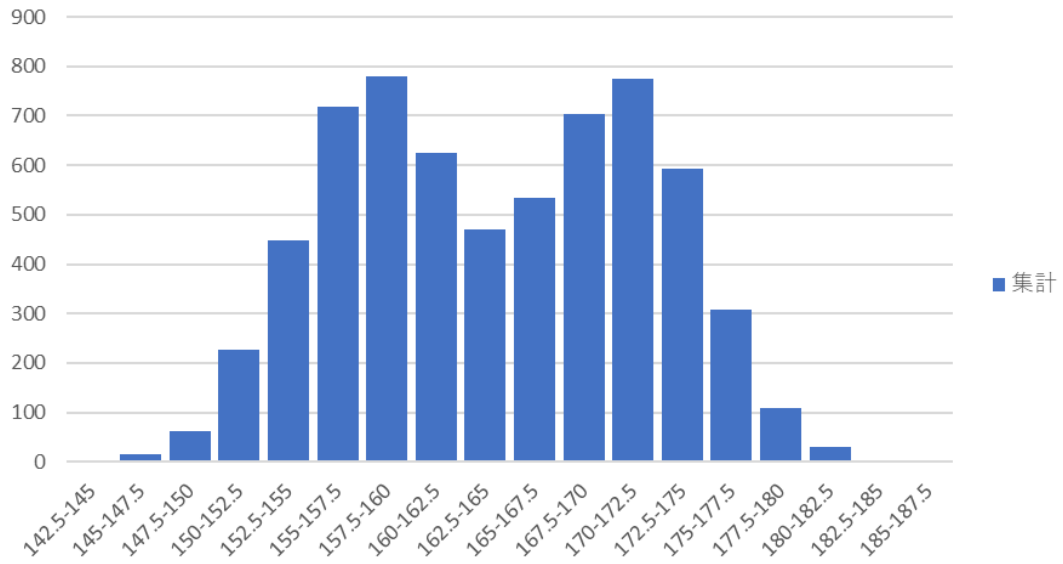
→160cm付近の身長が一般的（多い）とは限りません。

ヒストグラム

中央値の弱点

例：あるクラスの身長の中央値は160cm（平均値も160cm）

→160cm付近の身長が一般的（多い）とは限りません。

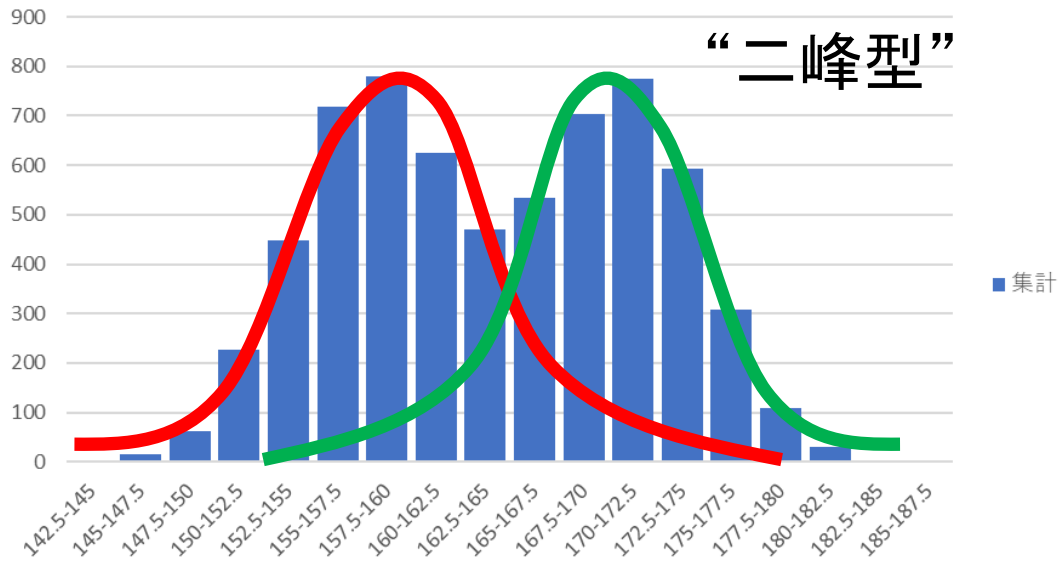


ヒストグラム

中央値の弱点

例：あるクラスの身長の中央値は160cm（平均値も160cm）

→160cm付近の身長が一般的（多い）とは限りません。

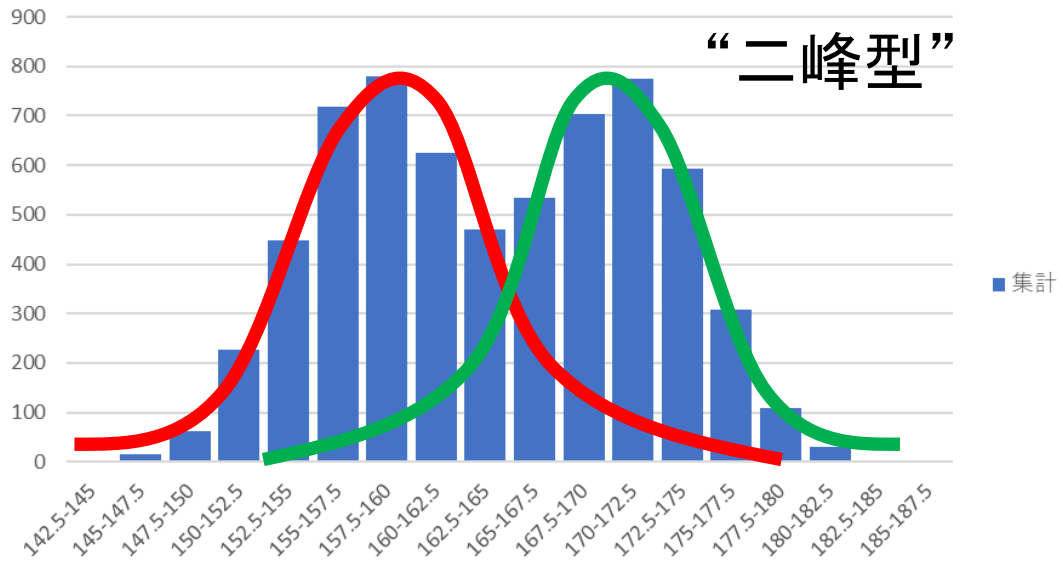


ヒストグラム

中央値の弱点

例：あるクラスの身長の中央値は160cm（平均値も160cm）

→160cm付近の身長が一般的（多い）とは限りません。

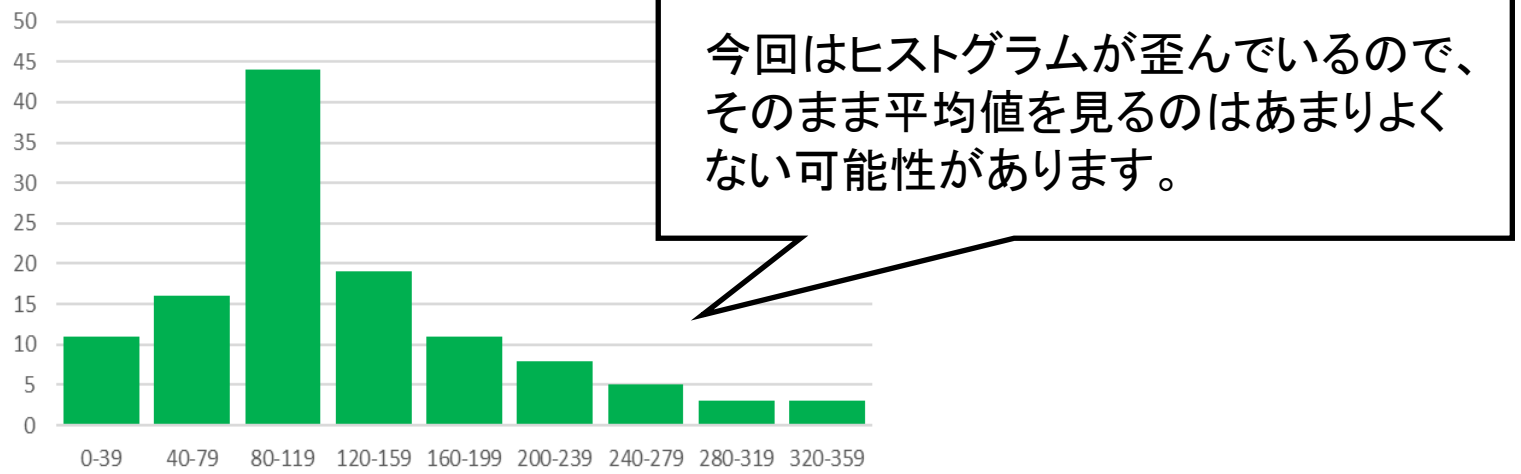


グラフを描くことでデータと特徴を
より正確に把握できる

ヒストグラムを見て

平均値だけでなく、他の指標もしっかり読み取ることが重要です。

グラフや他の指標と照らし合わせることで正しく把握する



今回はヒストグラムが歪んでいるので、そのまま平均値を見るのはあまりよくない可能性があります。

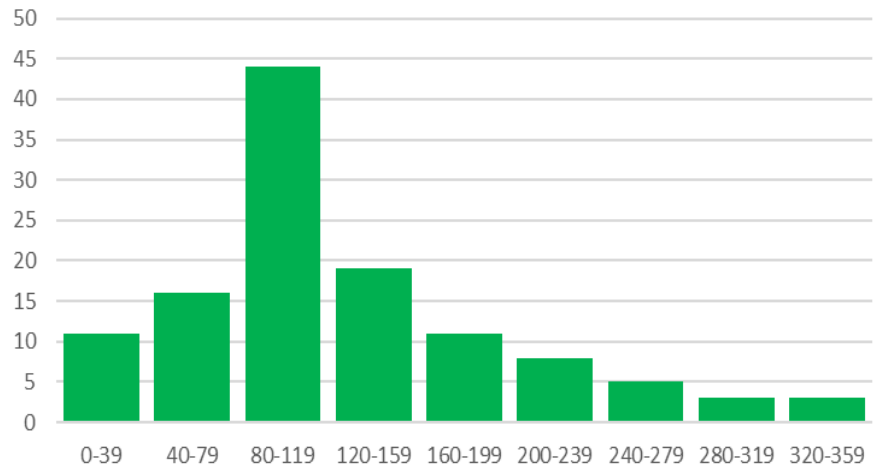
統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



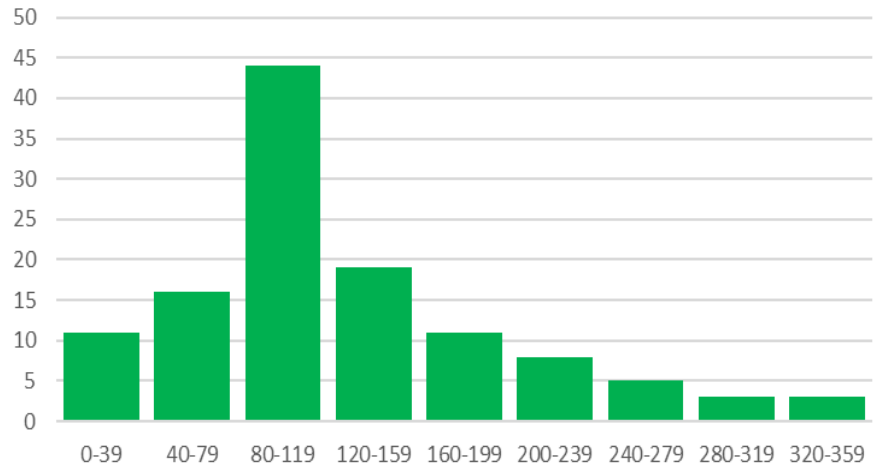
統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



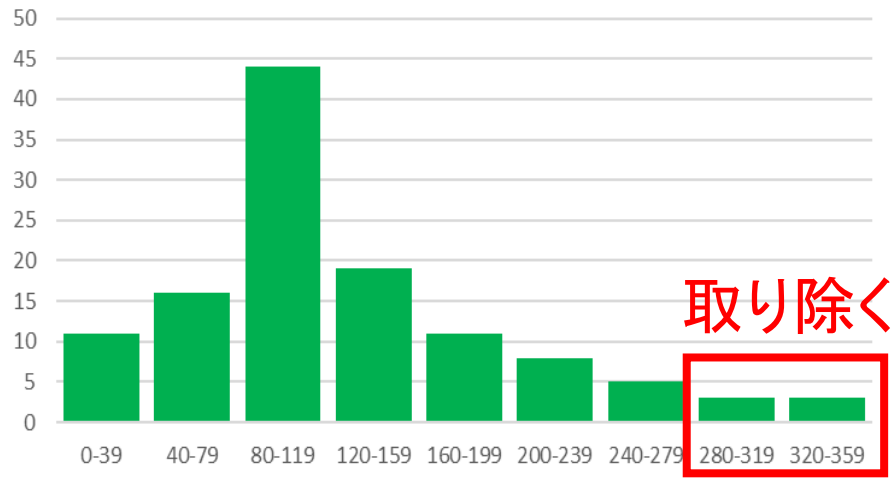
統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	109.5
２５％点	82
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



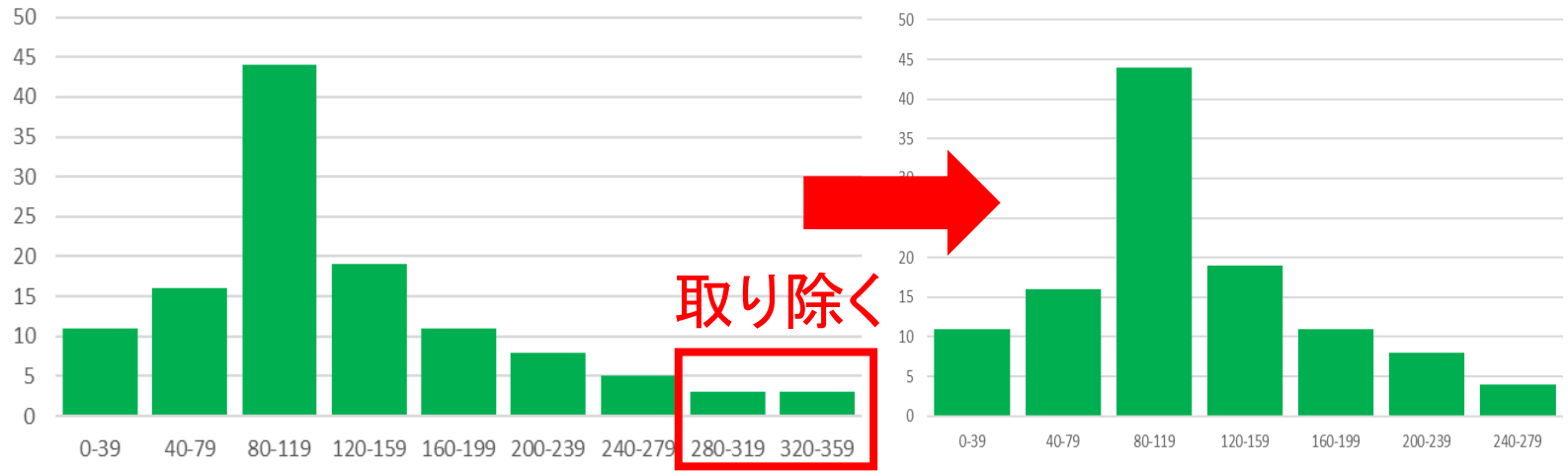
統計量	値
平均値	
最大値	
７５％点	
中央値	
２５％点	
最小値	
標準偏差	

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



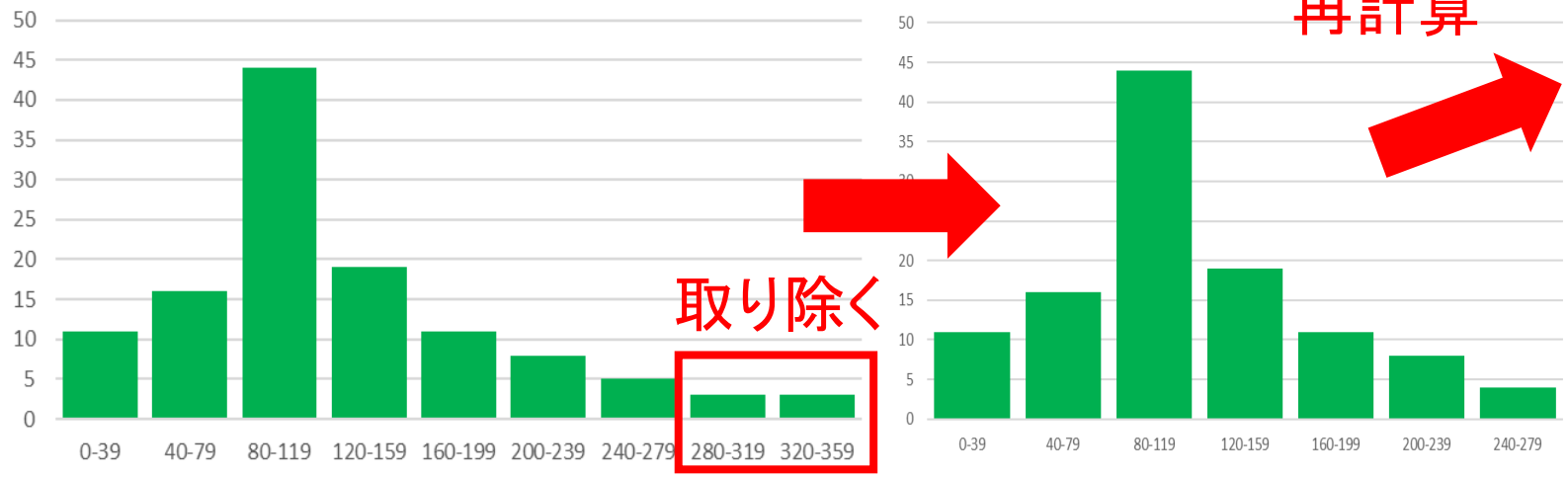
統計量	値
平均値	
最大値	
７５％点	
中央値	
２５％点	
最小値	
標準偏差	

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



統計量	値
平均値	114.6
最大値	261
７５％点	136
中央値	105
２５％点	81
最小値	3
標準偏差	57.4

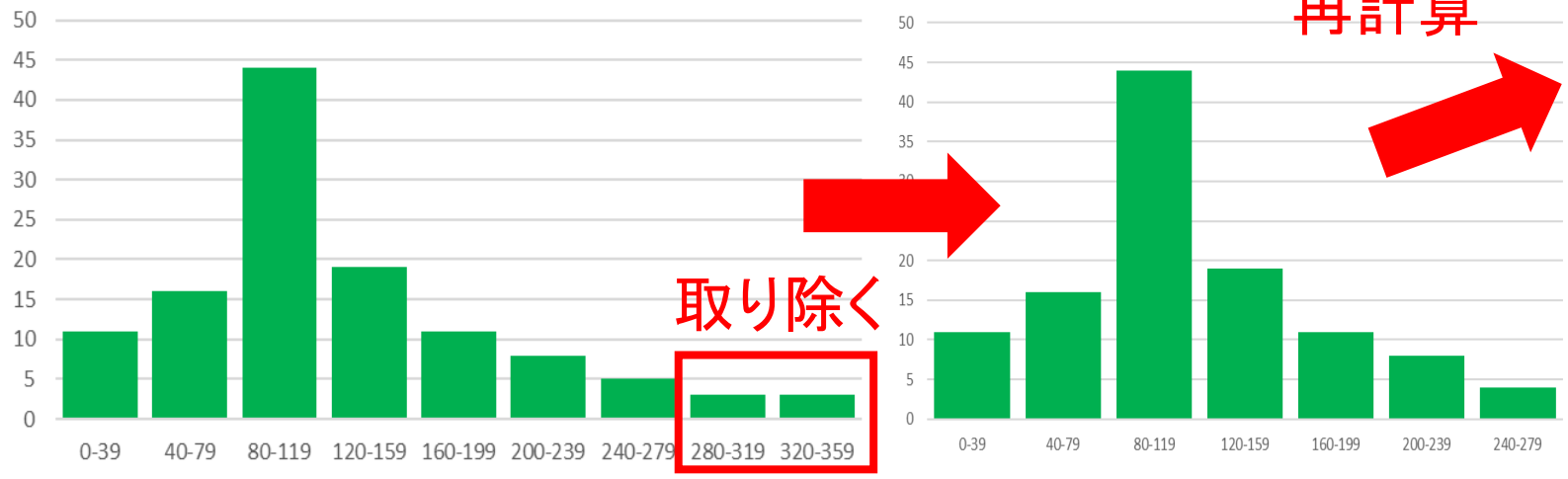
ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

「115+57=172個の在庫があればいいだろう」



統計量	値
平均値	**114.6**
最大値	261
７５％点	136
中央値	105
２５％点	81
最小値	3
標準偏差	**57.4**

参考：外れ値の判断

外れ値の判断方法の例：平均値と標準偏差を利用

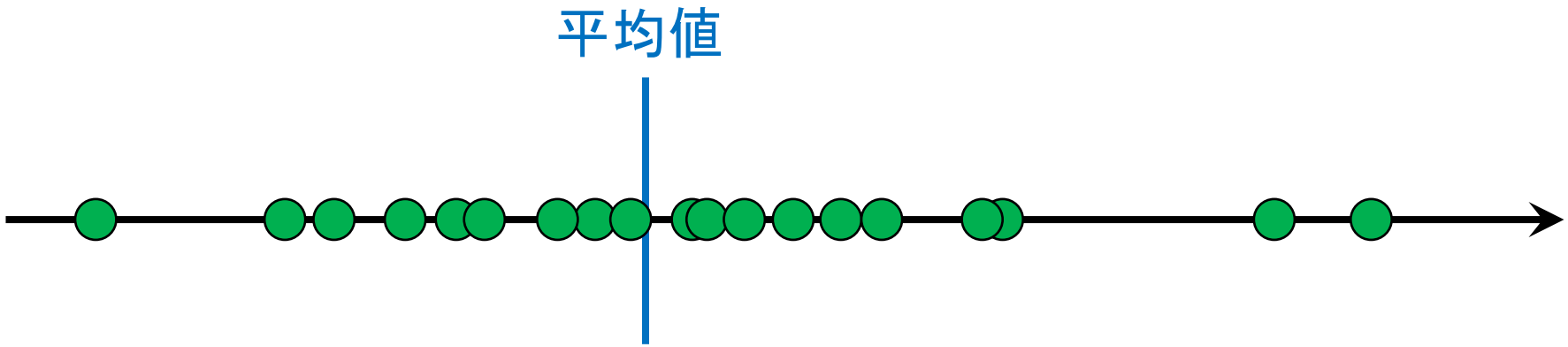
参考：外れ値の判断

外れ値の判断方法の例：平均値と標準偏差を利用



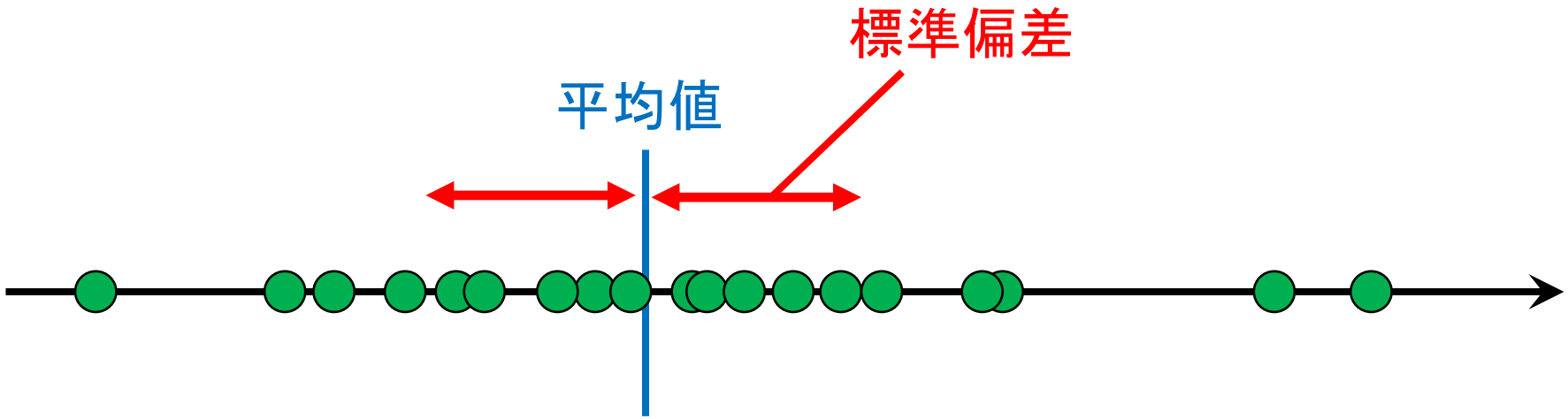
参考：外れ値の判断

外れ値の判断方法の例：平均値と標準偏差を利用



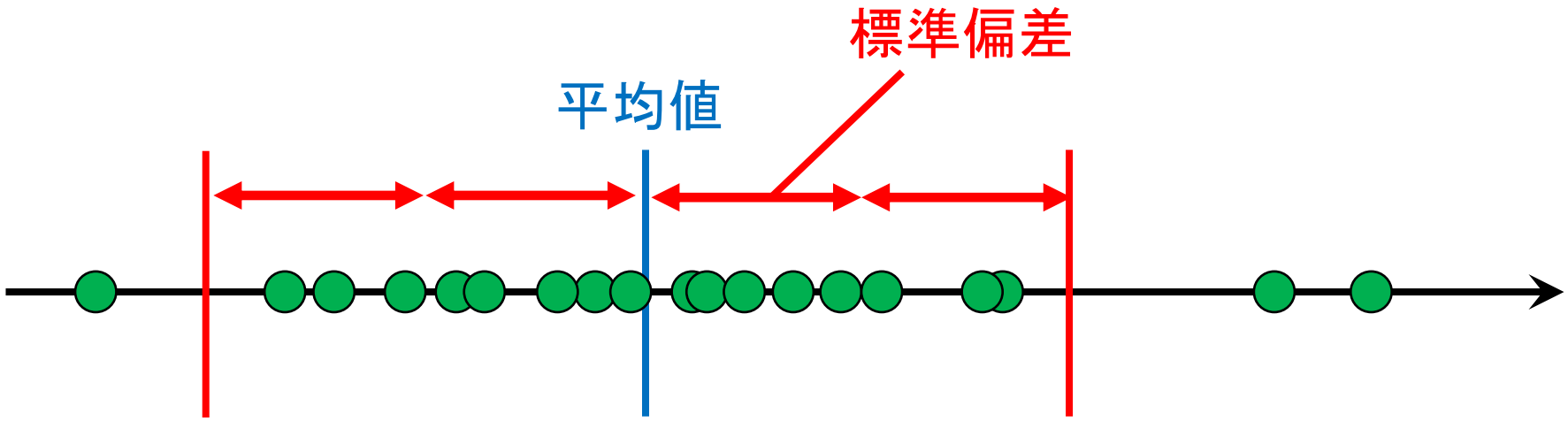
参考：外れ値の判断

外れ値の判断方法の例：平均値と標準偏差を利用



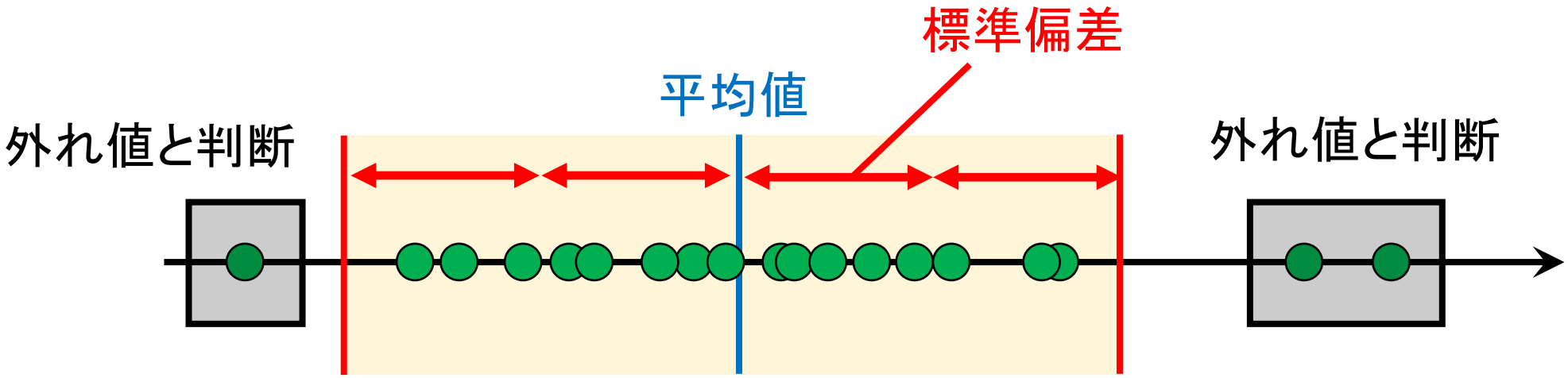
参考：外れ値の判断

外れ値の判断方法の例：平均値と標準偏差を利用



参考：外れ値の判断

外れ値の判断方法の例：平均値と標準偏差を利用



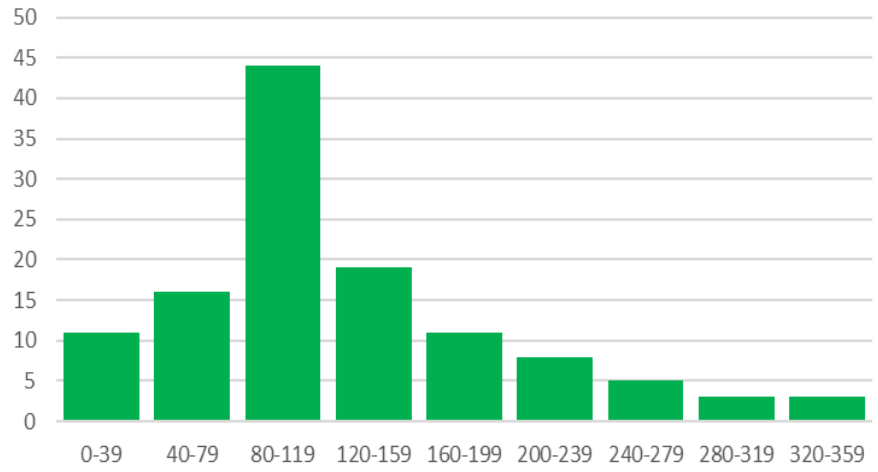
「平均値±２×標準偏差」の範囲外のデータを『外れ値』とみなす。

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



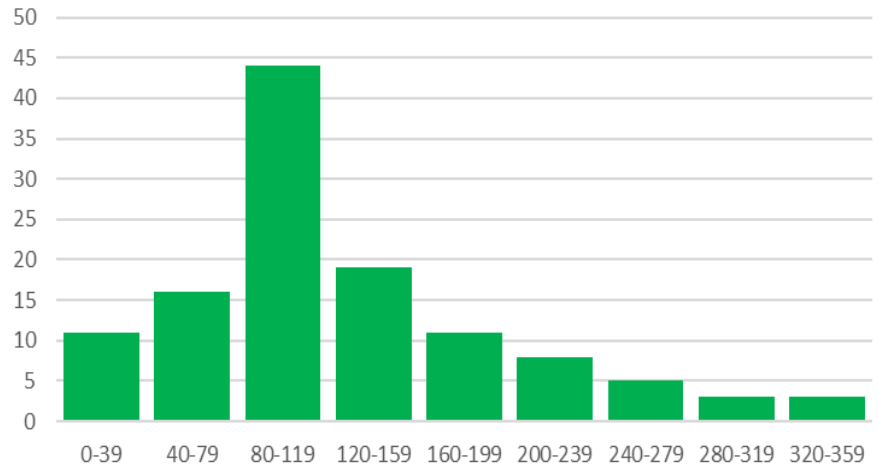
統計量	値
平均値	125.8
最大値	3
７５％点	161.5
中央値	109.5
２５％点	82
最小値	341
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



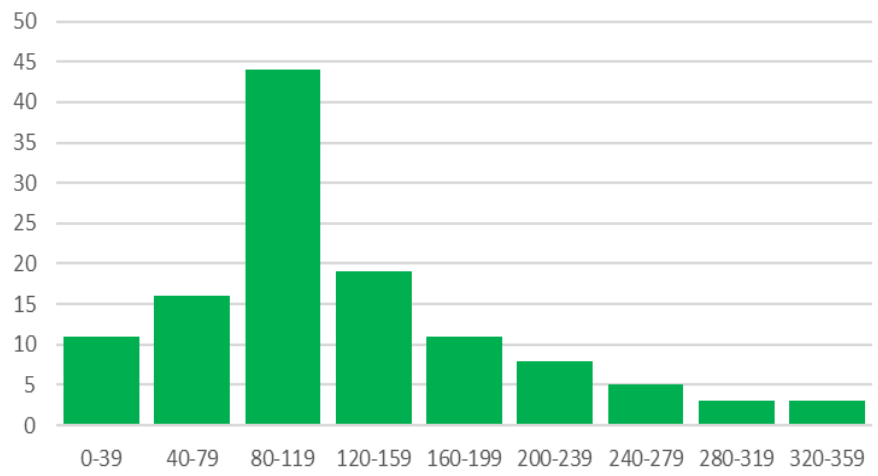
統計量	値
平均値	125.8
最大値	341
７５％点	161.5
中央値	**109.5**
２５％点	82
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る



統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0

平均値ではなく、中央値を基準にする場合、
ばらつきの度合いは四分位数を活用します。

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0



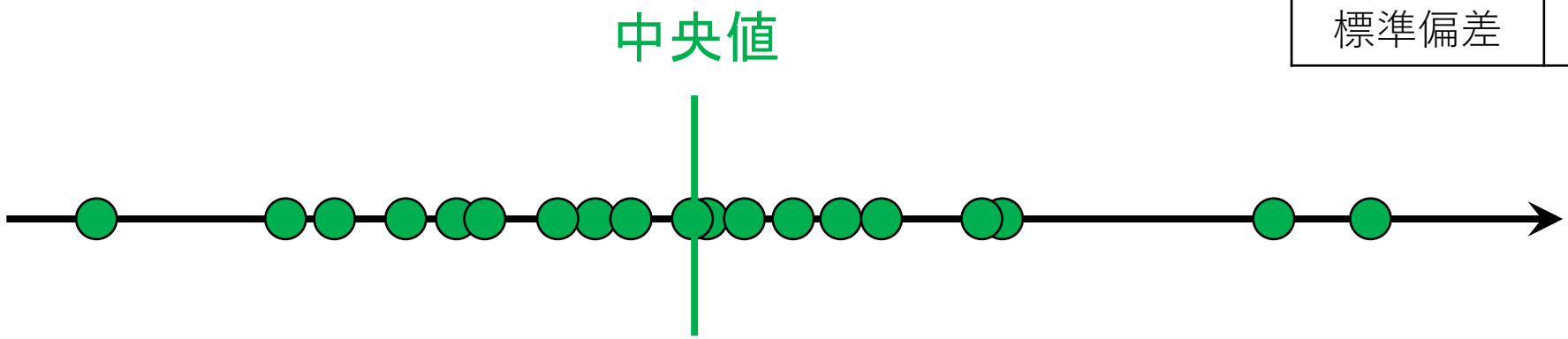
ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0



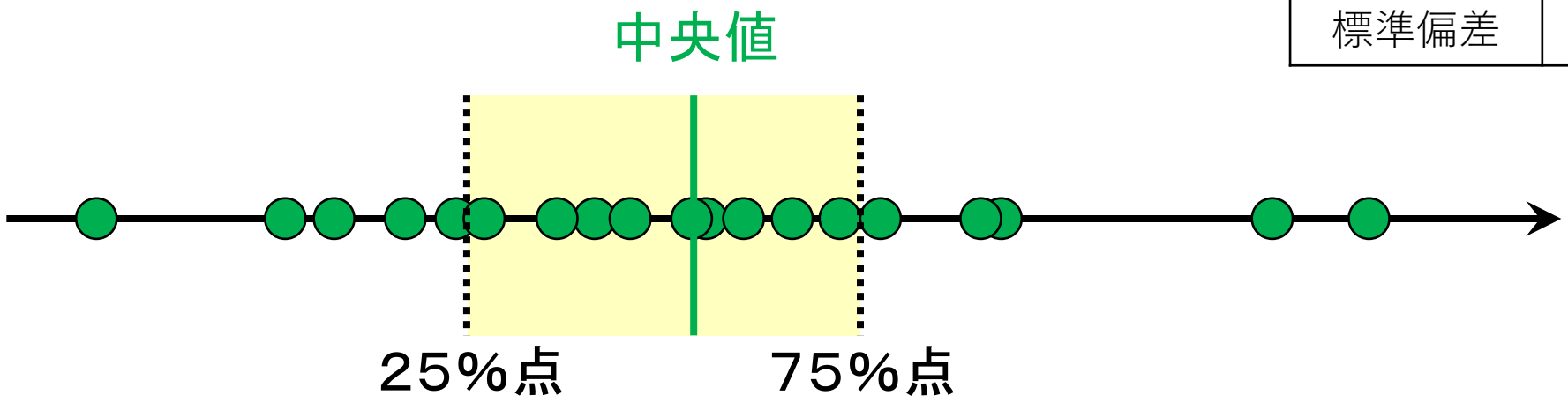
ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0



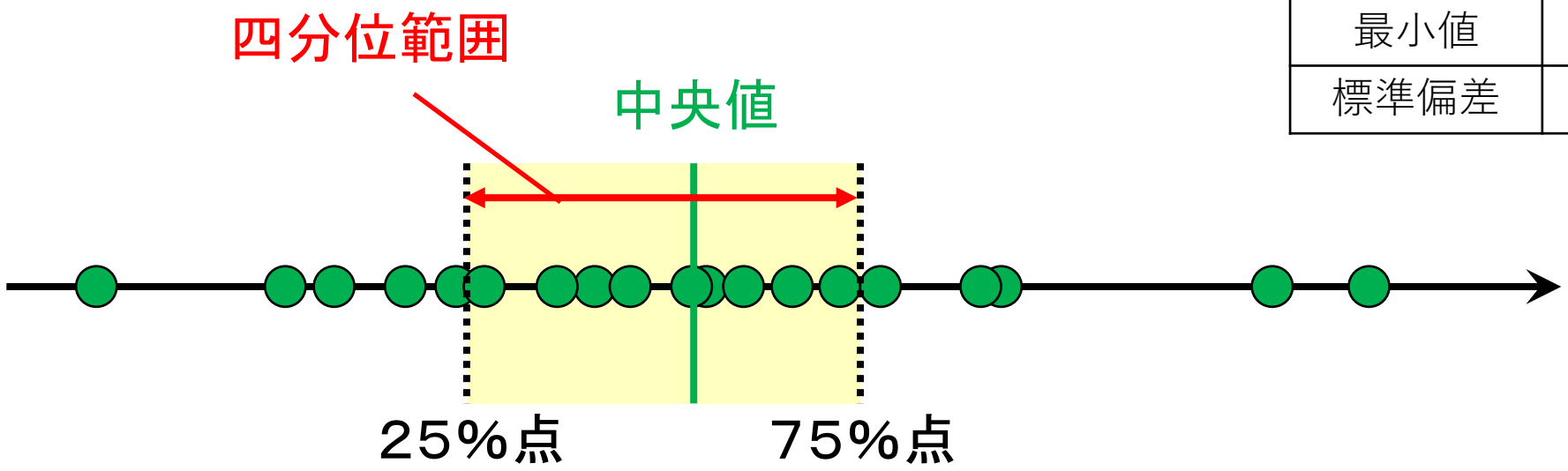
ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0



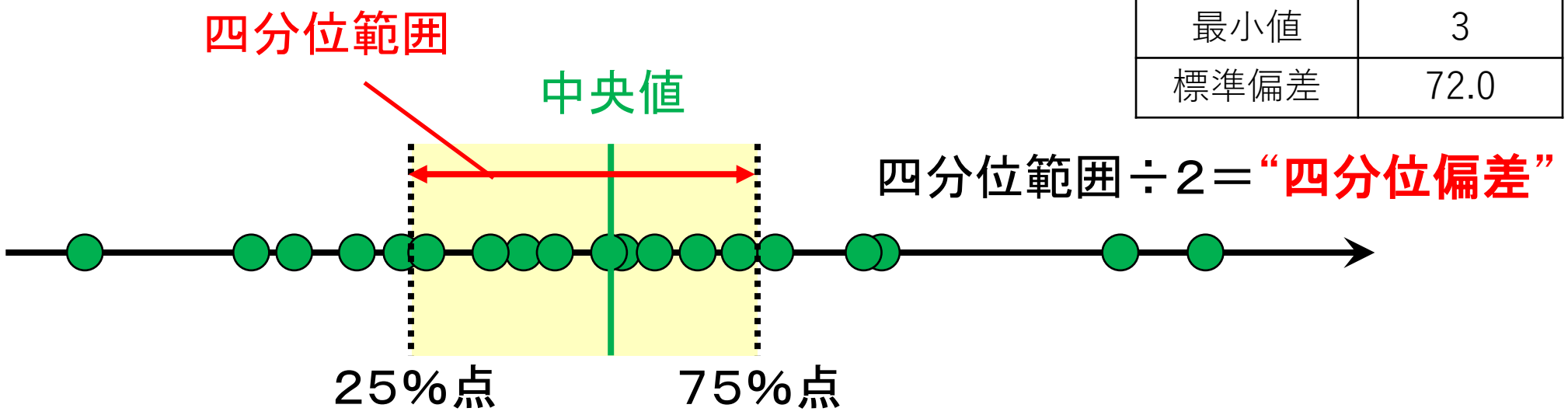
ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0



ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

四分位範囲＝161.5−82＝79.5

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

四分位範囲＝161.5−82＝79.5

四分位偏差＝79.5÷2=39.75

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

四分位範囲＝161.5−82＝79.5

四分位偏差＝79.5÷2=39.75

中央値＋四分位範囲＝109.5＋39.75 ＝ 149.25

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

② 中央値や四分位範囲を見る

四分位範囲＝161.5−82＝79.5

四分位偏差＝79.5÷2=39.75

中央値＋四分位範囲＝109.5＋39.75 ＝ 149.25

「150個在庫があればいいだろう」

統計量	値
平均値	125.8
最大値	341
７５％点	**161.5**
中央値	**109.5**
２５％点	**82**
最小値	3
標準偏差	72.0

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

②

四分

四分

中央

統計量	値
平均値	125.8
最大値	3
	61.5
	09.5
	82
	341
	2.0

指標（代表値）やグラフを使いながらデータの特徴を捉えていく

「150個在庫があればいいだろう」

ヒストグラムが歪んでいる場合

ヒストグラムが歪んでいる場合の対処法

① 外れ値を取り除き、指標を再計算する

②

四分

四分

中央

統計量	値
平均値	125.8
最大値	3
	61.5
	09.5
	82
	341
	2.0

指標（代表値）やグラフを使いながらデータの特徴を捉えていく

“記述統計”

「150個在庫があればいいだろう」

2つデータを比較する

2つの商品の特徴を比較してみましょう。

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94

2つデータを比較する

2つの商品の特徴を比較してみましょう。

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94



統計量	Aの代表値	Bの代表値
平均値	125.8	124.2
最大値	341	227
７５％点	161.5	155.3
中央値	109.5	124
２５％点	82	95.8
最小値	3	11
標準偏差	72.0	42.2

2つデータを比較する

2つの商品の特徴を比較してみましょう。

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94



統計量	Aの代表値	Bの代表値
平均値	**125.8**	**124.2**
最大値	341	227
75%点	161.5	155.3
中央値	109.5	124
25%点	82	95.8
最小値	3	11
標準偏差	**72.0**	**42.2**

AとBの平均値には差がないですが、Aの方がBに比べてばらつきが大きいことがわかります。

2つデータを比較する

2つの商品の特徴を比較してみましょう。

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94



統計量	Aの代表値	Bの代表値
平均値	**125.8**	**124.2**
最大値	341	227
７５％点	161.5	155.3
中央値	109.5	124
２５％点	82	95.8
最小値	3	11
標準偏差	**72.0**	**42.2**

AとBの平均値には差がないですが、Aの方がBに比べてばらつきが大きいことがわかります。

「Bの方が安定して売れているのではないか」

なぜAの商品はばらついているのか？

では、なぜAの商品はばらついているのでしょうか？

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94

統計量	Aの代表値	Bの代表値
平均値	125.8	124.2
最大値	341	227
７５％点	161.5	155.3
中央値	109.5	124
２５％点	82	95.8
最小値	3	11
標準偏差	**72.0**	42.2

なぜAの商品はばらついているのか？

では、なぜAの商品はばらついているのでしょうか？

データの種類や情報を増やし、『ばらつきを生みだしている理由』を探ってみましょう。

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94

統計量	Aの代表値	Bの代表値
平均値	125.8	124.2
最大値	341	227
７５％点	161.5	155.3
中央値	109.5	124
２５％点	82	95.8
最小値	3	11
標準偏差	**72.0**	42.2

なぜAの商品はばらついているのか？

では、なぜAの商品はばらついているのでしょうか？

データの種類や情報を増やし、『ばらつきを生みだしている理由』を探ってみましょう。

日付	Aの販売量	Bの販売量
2021/7/1	106	203
2021/7/2	113	187
2021/7/3	201	130
2021/7/4	320	124
2021/7/5	128	162
2021/7/6	58	139
2021/7/7	91	128
2021/7/8	45	82
2021/7/9	126	106
2021/7/10	159	94

＋

曜日	平均気温	天気
木曜日	20.9	雨
金曜日	21.2	雨
土曜日	22.2	雨
日曜日	20.1	雨
月曜日	22.1	曇
火曜日	24.7	曇
水曜日	25.8	曇
木曜日	23.7	雨
金曜日	23.5	雨
土曜日	27	晴れ

統計量	Aの代表値	Bの代表値
平均値	125.8	124.2
最大値	341	227
７５％点	161.5	155.3
中央値	109.5	124
２５％点	82	95.8
最小値	3	11
標準偏差	**72.0**	42.2

なぜAの商品はばらついているのか？

例：曜日ごとのAの販売量を集計してみましょう。

日付	Aの販売量	Bの販売量	曜日	平均気温	天気
2021/7/1	106	203	木曜日	20.9	雨
2021/7/2	113	187	金曜日	21.2	雨
2021/7/3	201	130	土曜日	22.2	雨
2021/7/4	320	124	日曜日	20.1	雨
2021/7/5	128	162	月曜日	22.1	曇
2021/7/6	58	139	火曜日	24.7	曇
2021/7/7	91	128	水曜日	25.8	曇
2021/7/8	45	82	木曜日	23.7	雨
2021/7/9	126	106	金曜日	23.5	雨
2021/7/10	159	94	土曜日	27	晴れ

なぜAの商品はばらついているのか？

例：曜日ごとのAの販売量を集計してみましょう。

日付	Aの販売量	Bの販売量	曜日	平均気温	天気
2021/7/1	106	203	木曜日	20.9	雨
2021/7/2	113	187	金曜日	21.2	雨
2021/7/3	201	130	土曜日	22.2	雨
2021/7/4	320	124	日曜日	20.1	雨
2021/7/5	128	162	月曜日	22.1	曇
2021/7/6	58	139	火曜日	24.7	曇
2021/7/7	91	128	水曜日	25.8	曇
2021/7/8	45	82	木曜日	23.7	雨
2021/7/9	126	106	金曜日	23.5	雨
2021/7/10	159	94	土曜日	27	晴れ

なぜAの商品はばらついているのか？

例：曜日ごとのAの販売量を集計してみましょう。

日付	Aの販売量	Bの販売量	曜日	平均気温	天気
2021/7/1	106	203	木曜日	20.9	雨
2021/7/2	113	187	金曜日	21.2	雨
2021/7/3	201	130	土曜日	22.2	雨
2021/7/4	320	124	日曜日	20.1	雨
2021/7/5	128	162	月曜日	22.1	曇
2021/7/6	58	139	火曜日	24.7	曇
2021/7/7	91	128	水曜日	25.8	曇
2021/7/8	45	82	木曜日	23.7	雨
2021/7/9	126	106	金曜日	23.5	雨
2021/7/10	159	94	土曜日	27	晴れ

曜日	商品Aの平均売上
月曜日	84.4
火曜日	95.5
水曜日	82.9
木曜日	88.2
金曜日	92.8
土曜日	177.1
日曜日	261.8

なぜAの商品はばらついているのか？

例：曜日ごとのAの販売量を集計してみましょう。

日付	Aの販売量	Bの販売量	曜日	平均気温	天気
2021/7/1	106	203	木曜日	20.9	雨
2021/7/2	113	187	金曜日	21.2	雨
2021/7/3	201	130	土曜日	22.2	雨
2021/7/4	320	124	日曜日	20.1	雨
2021/7/5	128	162	月曜日	22.1	曇
2021/7/6	58	139	火曜日	24.7	曇
2021/7/7	91	128	水曜日	25.8	曇
2021/7/8	45	82	木曜日	23.7	雨
2021/7/9	126	106	金曜日	23.5	雨
2021/7/10	159	94	土曜日	27	晴れ

週末の売上が多く、曜日によって差があることから、ばらつきが生じたと考えられます。

曜日	商品Aの平均売上
月曜日	84.4
火曜日	95.5
水曜日	82.9
木曜日	88.2
金曜日	92.8
土曜日	177.1
日曜日	261.8

なぜAの商品はばらついているのか？

例：曜日ごとのAの販売量を集計してみましょう。

週末の売上が多く、曜日によって差があることから、ばらつきが生じたと考えられます。

このような複数のデータを関数のみで集計するのは大変です。

日付					
2021/7/8	45	82	木曜日	23.7	雨
2021/7/9	126	106	金曜日	23.5	雨
2021/7/10	159	94	土曜日	27	晴れ

売上
177.1
261.8

なぜAの商品はばらついているのか？

例：曜日ごとのAの販売量を集計してみましょう。

週末の売上が多く、曜日によって差があることから、ばらつきが生じたと考えられます。

このような複数のデータを関数のみで集計するのは大変です。

⇩

Excelの「ピボットテーブル」という機能を使って集計してみましょう！

日付					
2021/7/8	45	82	木曜日	23.7	雨
2021/7/9	126	106	金曜日	23.5	雨
2021/7/10	159	94	土曜日	27	晴れ

売上
177.1
261.8

講義スケジュール

章	内容	目標
第1章	統計学とは何か？	集計と可視化によるデータの特徴を把握する重要性を理解する
第2章	情報収集とデータ把握	ピボットテーブルの使い方、オープンデータのDL方法～分析方法を理解する
第3章	データ分析に必要な思考力	データを用いた意思決定の手順について理解する

講義スケジュール

章	内容	目標
第1章	統計学とは何か？	集計と可視化によるデータの特徴を把握する重要性を理解する
第2章	情報収集とデータ把握	ピボットテーブルの使い方、オープンデータのDL方法～分析方法を理解する
第3章	データ分析に必要な思考力	データを用いた意思決定の手順について理解する