

Excelで学ぶデータ分析

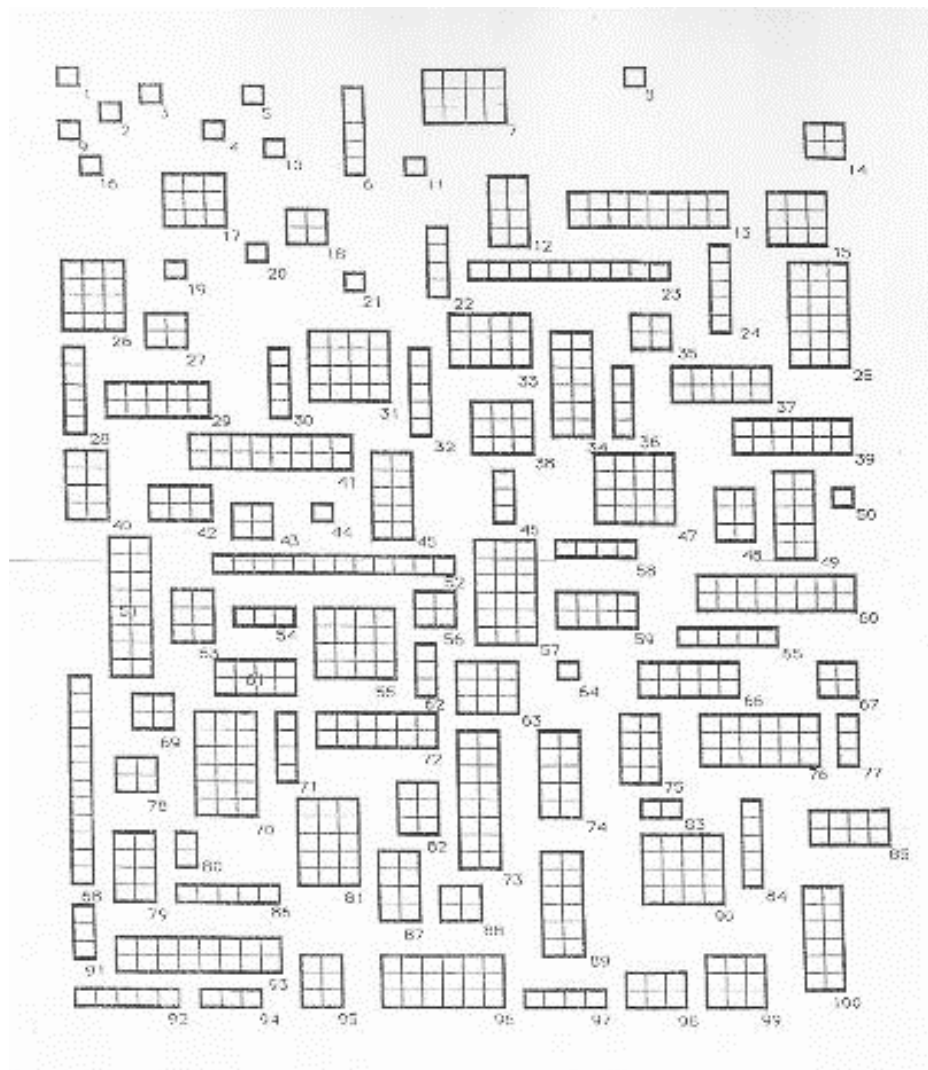
第2回

「無作為抽出・確率」



和から株式会社

演習 1



(1) 各図形が含む正方形の数の平均値は？

統計学の分類

統計学

```
graph TD; A[統計学] --- B[記述統計学]; A --- C[推測統計学];
```

記述統計学

- ✓ データの整理し、データの特徴をできるだけ簡潔に表すことが目的
- ✓ 手法としては数値や表、グラフなどを用いてデータの特徴を捉える
- ✓ 例：国勢調査、営業成績

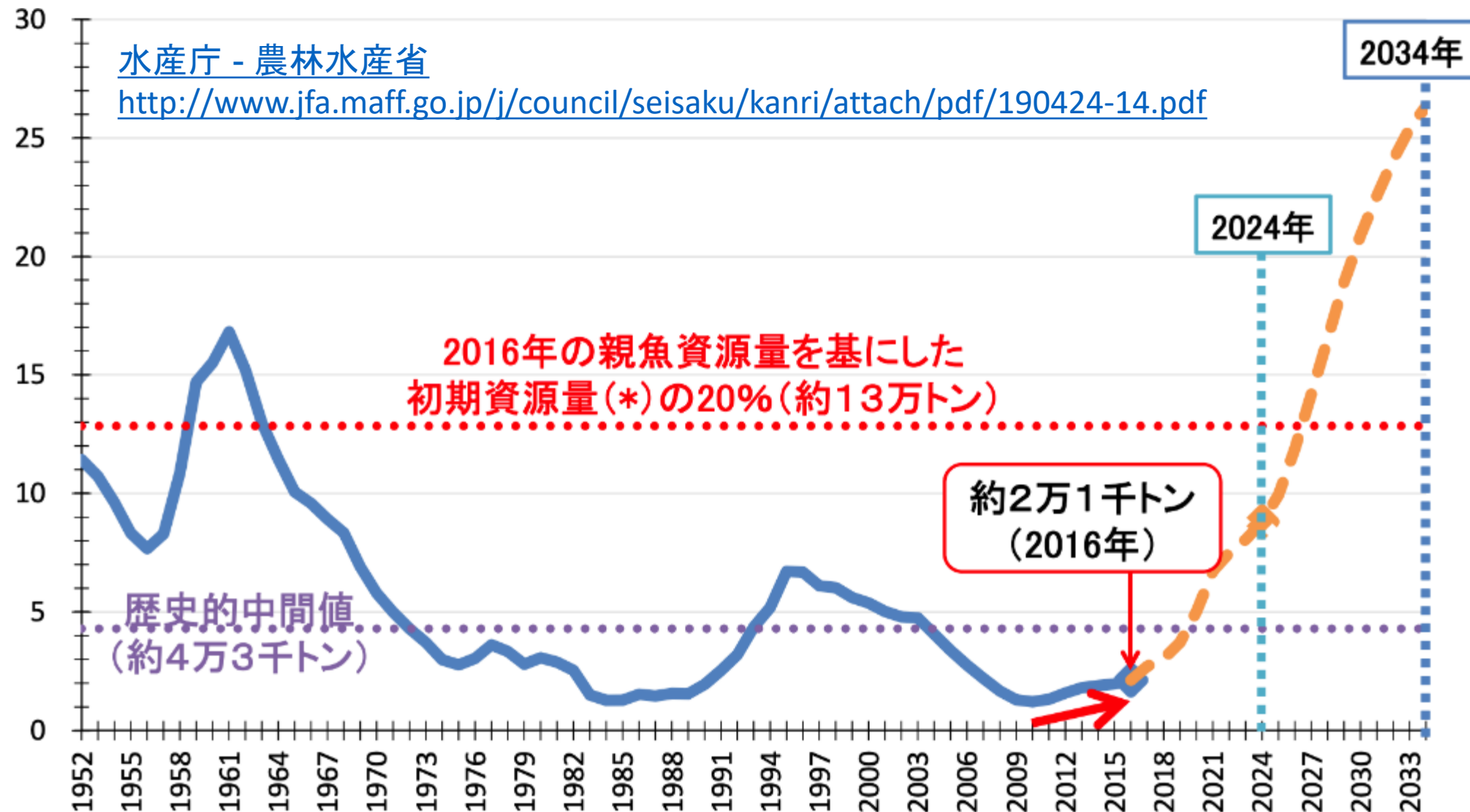
推測統計学

- ✓ サンプルデータ（標本）から全体（母集団）の状況を推測することが目的
- ✓ 推測統計学には2つの手法がある
 - 推定
 - 検定

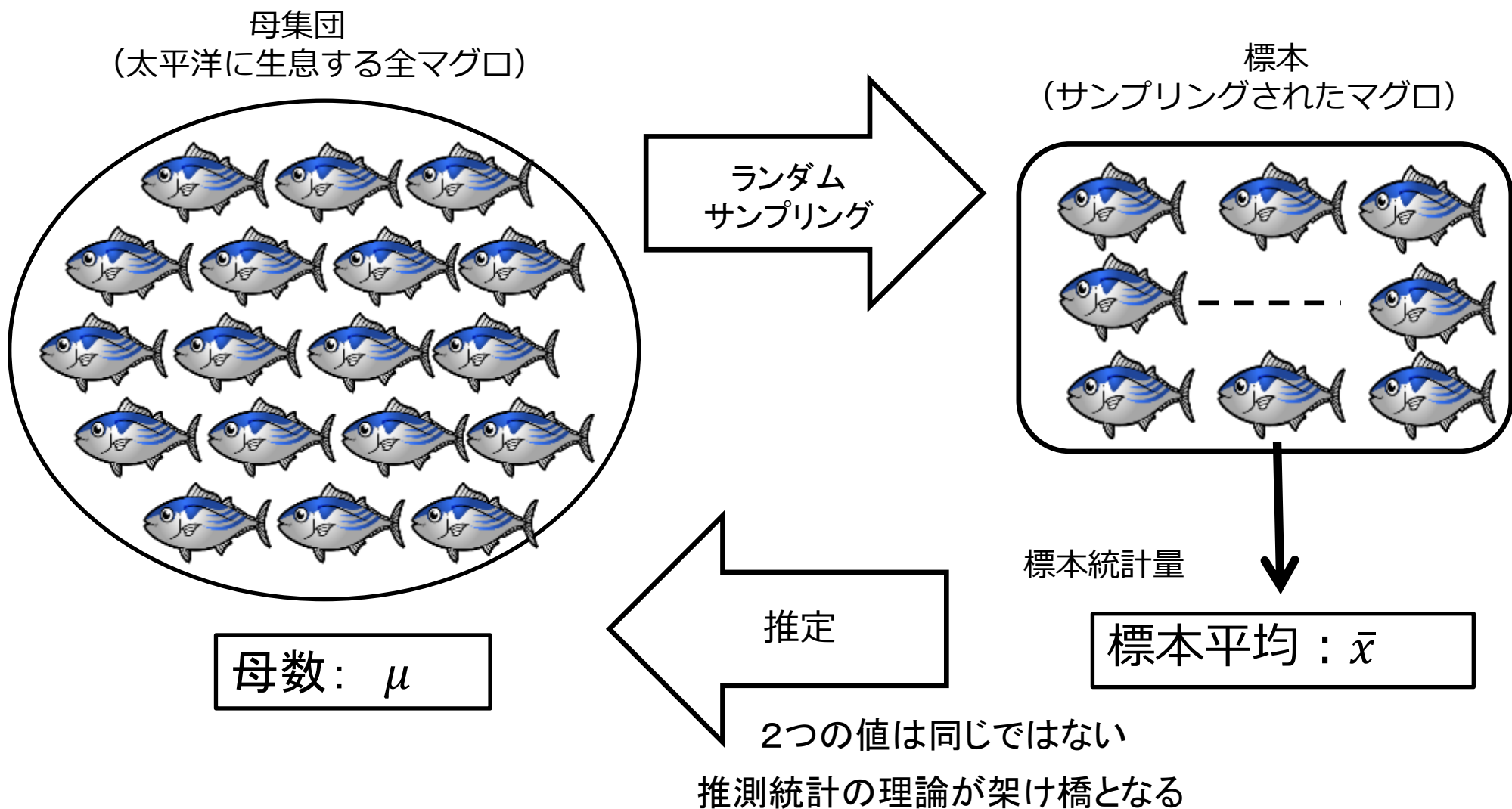
太平洋クロマグロの資源量推定

水産庁 - 農林水産省

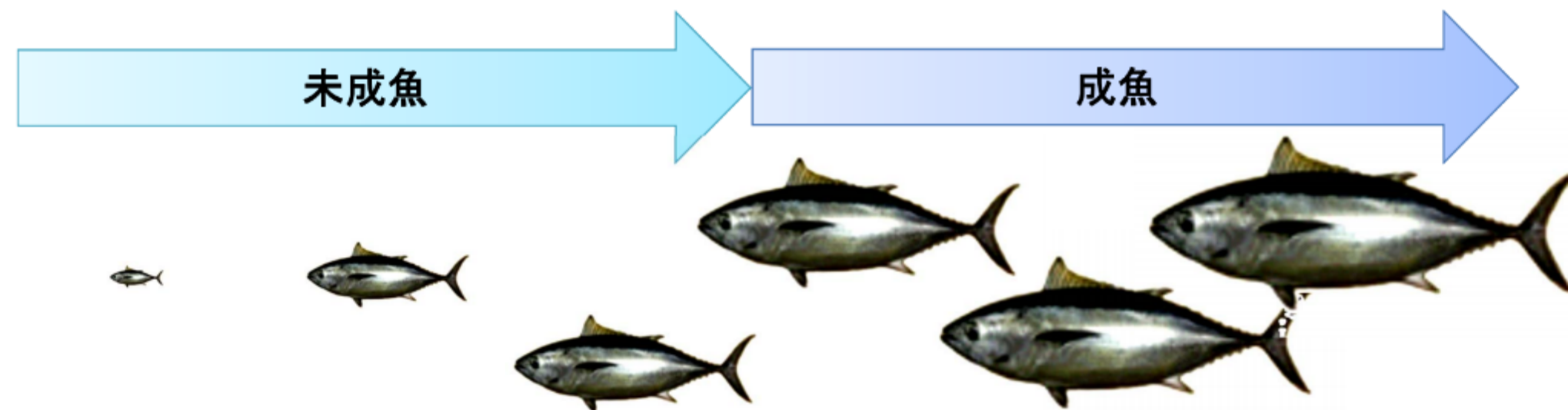
<http://www.jfa.maff.go.jp/j/council/seisaku/kanri/attach/pdf/190424-14.pdf>



太平洋に生息するマグロの平均体重は？



太平洋クロマグロの成長



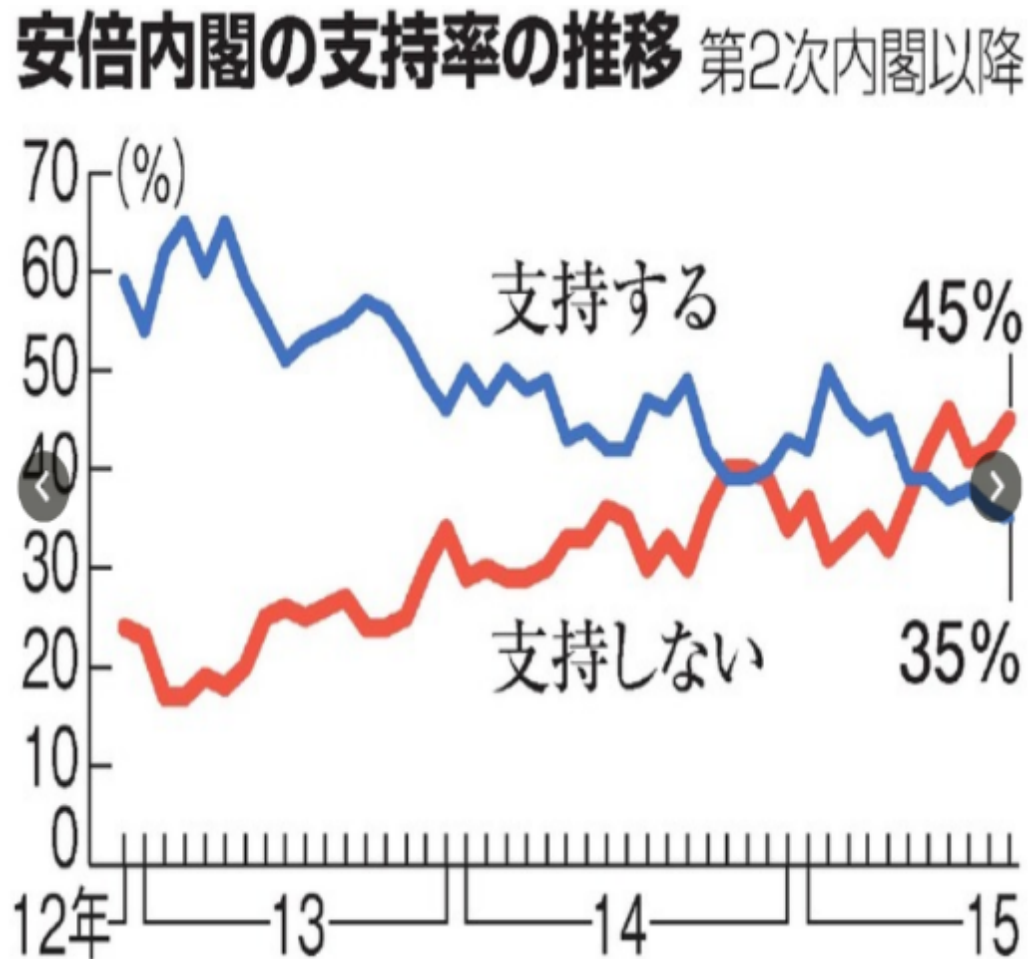
0歳	1歳	2歳	3歳	4歳	5歳
19. 1cm	58. 6cm	91. 4cm	118. 6cm	141. 1cm	159. 7cm
0. 2kg	4. 4kg	16. 1kg	34. 5kg	58. 4kg	85. 2kg
漁獲の対象 となり始める			全体の 20% が成熟	全体の 50% が成熟	全体の 100% が成熟

標本調査の事例

読売新聞



朝日新聞



検定

オバマ大統領が簡単なテストで、6000万ドルもの収益を上げた方法



検定の応用

Join ABCSPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

Sign up +

Type A

Join ABCSPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

100% privacy. We will never spam you !

Sign up +

Type B

ABテスト

パターン1（画像、オリジナル）：「Obama」の旗に囲まれる柔らかな

パターン2（画像）：家族と一緒に写っている写真

パターン3（画像）：正面からアップで撮影した凛々しい表情の写真



ABテスト

パターン1（オリジナル） : SIGN UP「会員登録」

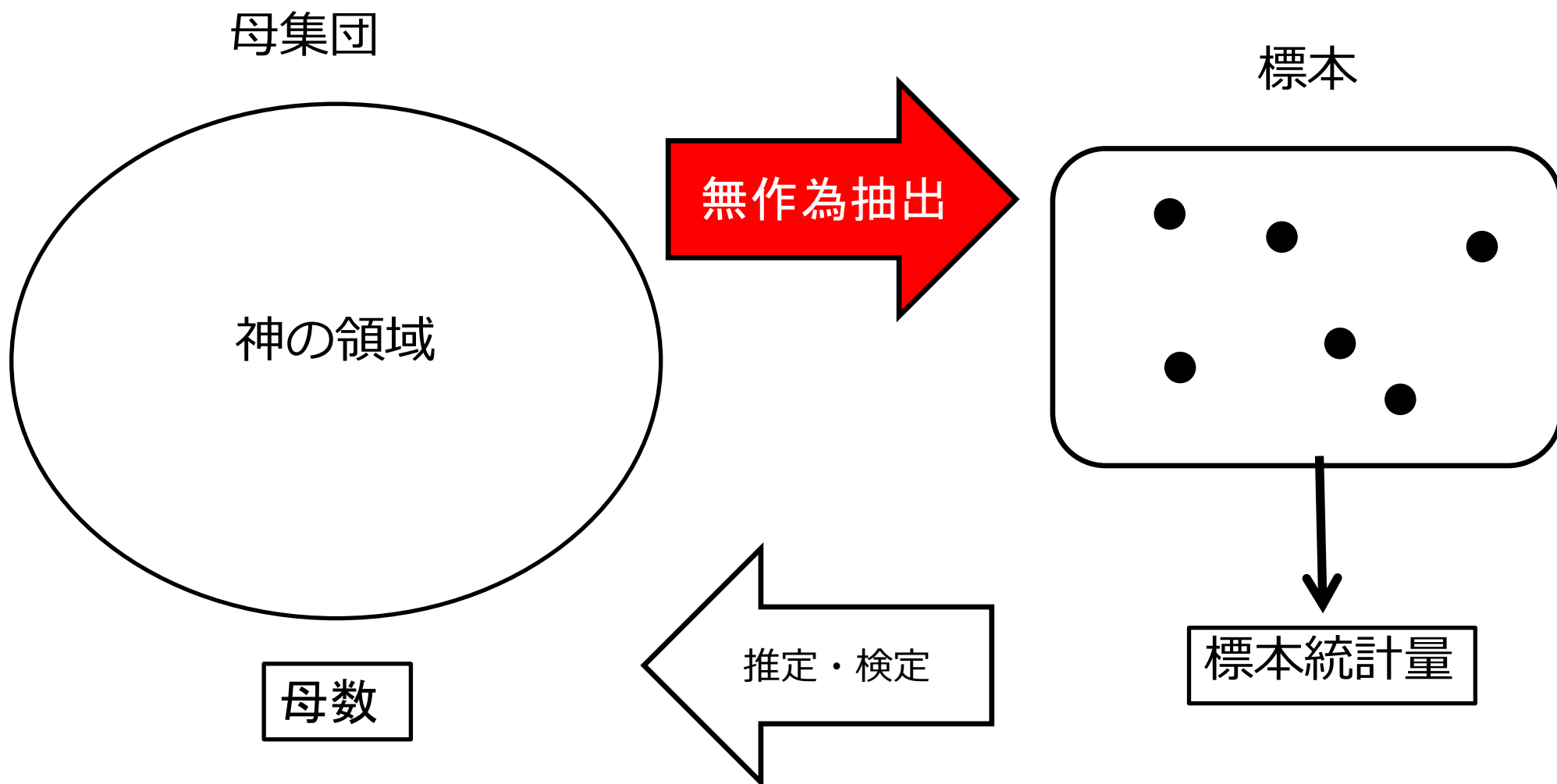
パターン2 : SIGN UP NOW「今すぐ会員登録」

パターン3 : JOIN US NOW「今すぐ参加する」

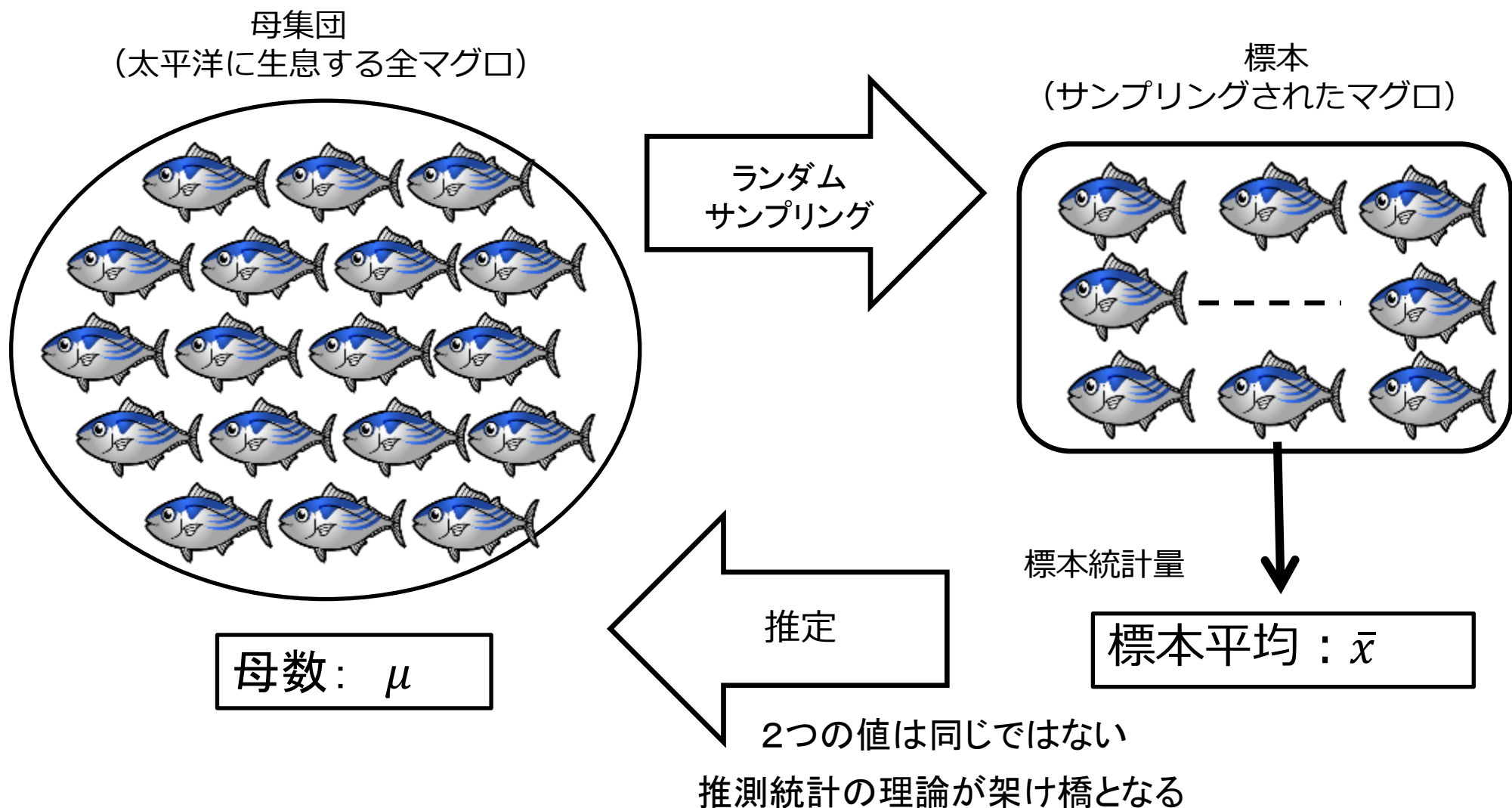
パターン4 : LEARN MORE「もっと詳しく」

A red rectangular button with rounded corners and a black border. The text "SIGN UP" is centered in white, bold, uppercase letters.A red rectangular button with rounded corners and a black border. The text "SIGN UP NOW" is centered in white, bold, uppercase letters.A red rectangular button with rounded corners and a black border. The text "JOIN US NOW" is centered in white, bold, uppercase letters.A red rectangular button with rounded corners and a black border. The text "LEARN MORE" is centered in white, bold, uppercase letters.

推定・検定と無作為抽出



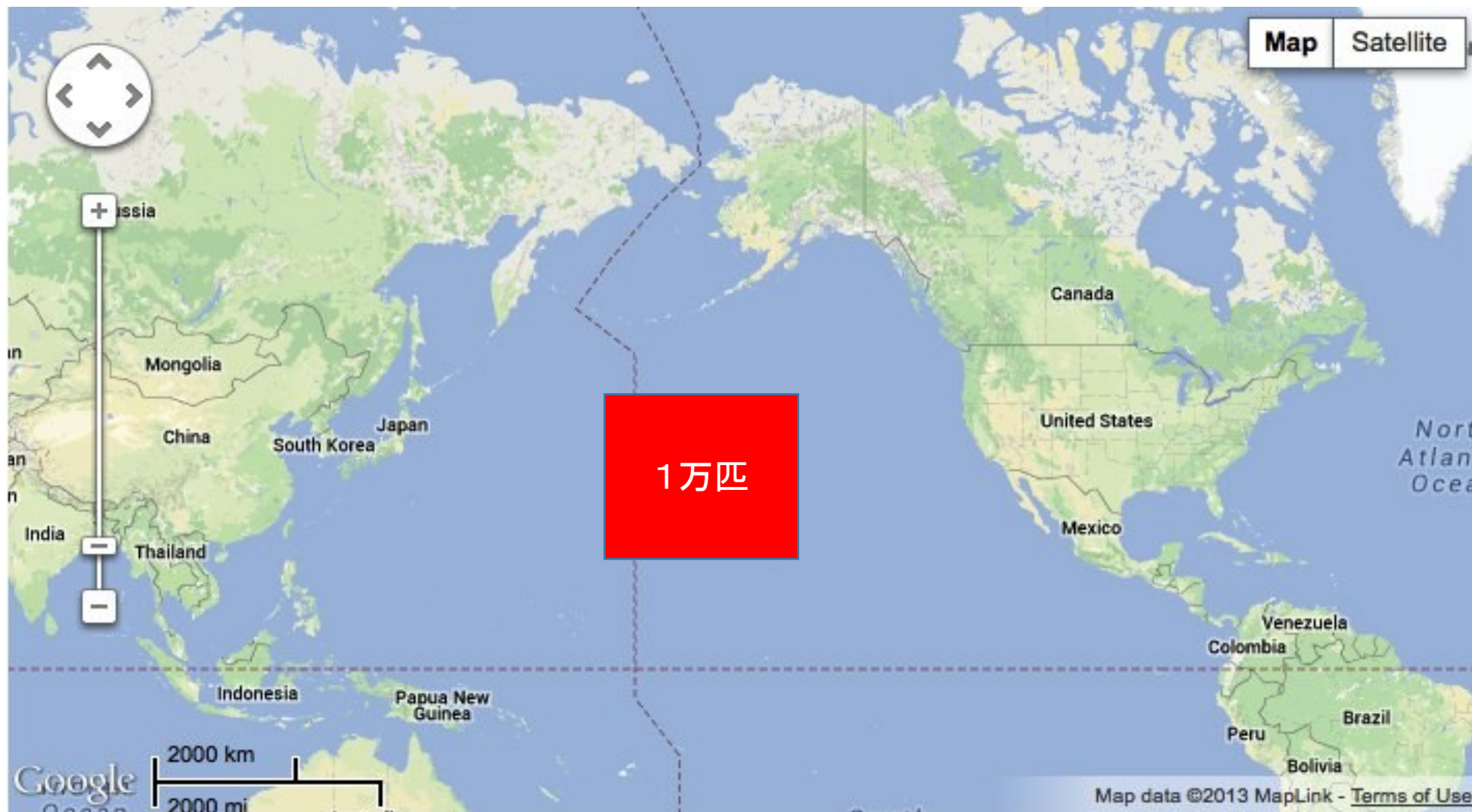
太平洋に生息するマグロの平均体重は？



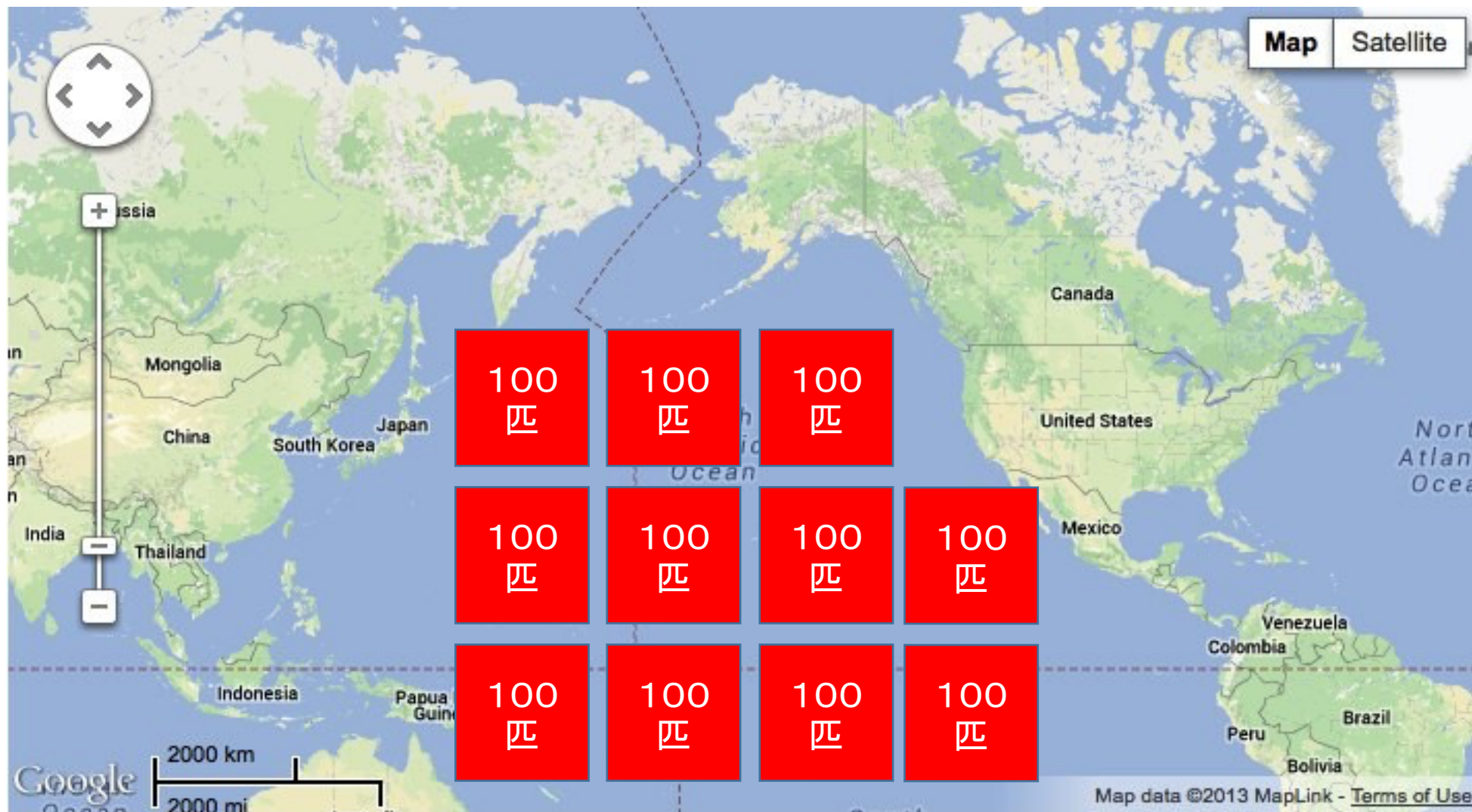
無作為抽出の重要性



無作為抽出の重要性



無作為抽出の重要性



アメリカ大統領選挙の番狂わせ

1936年のアメリカ大統領選挙



民主党
フランクリン・ルーズベルト



共和党
アルフレッド・ランドン

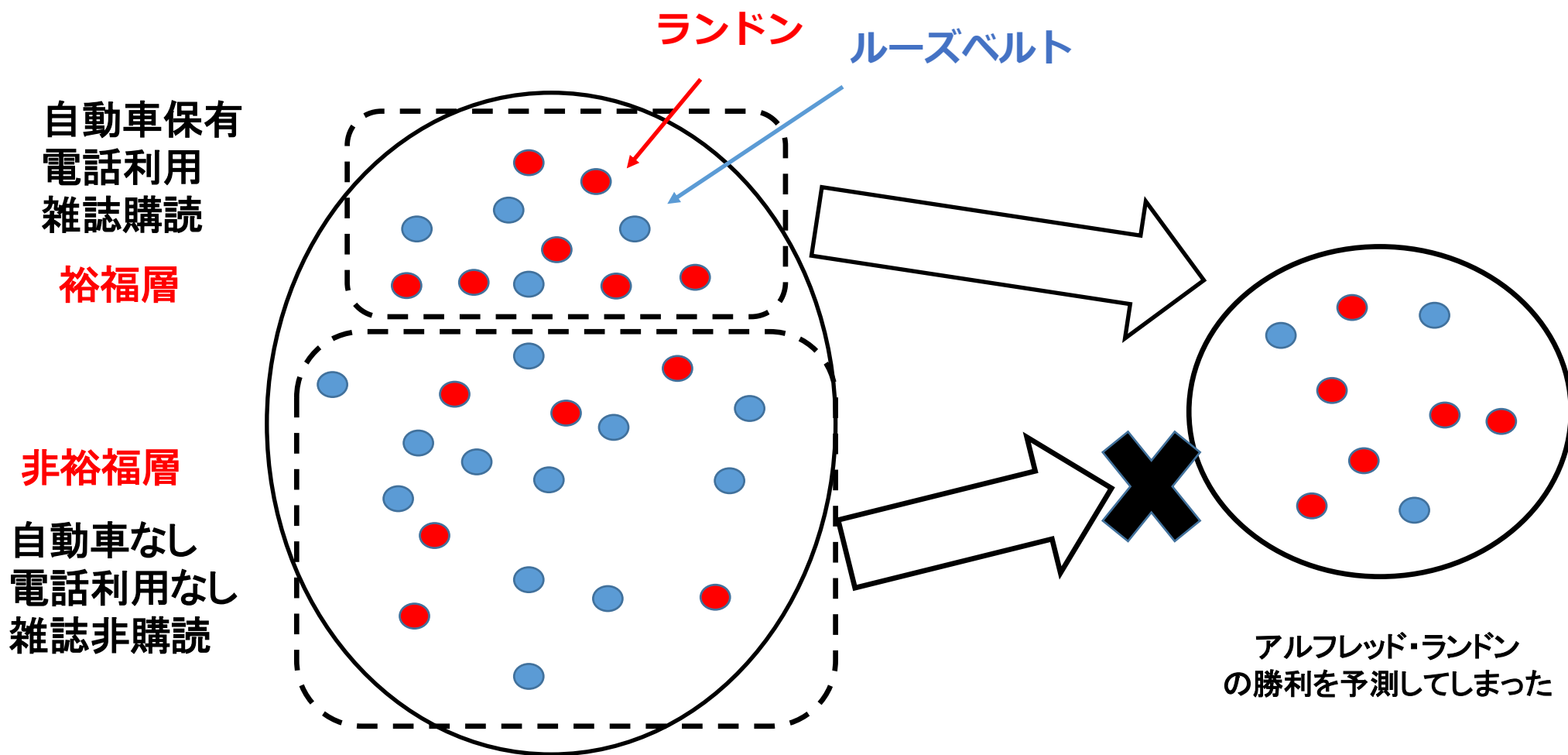
リテラリー・ダイジェスト社

200万人を対象に調査を行い、ランドンが57%の得票を得て当選すると予想

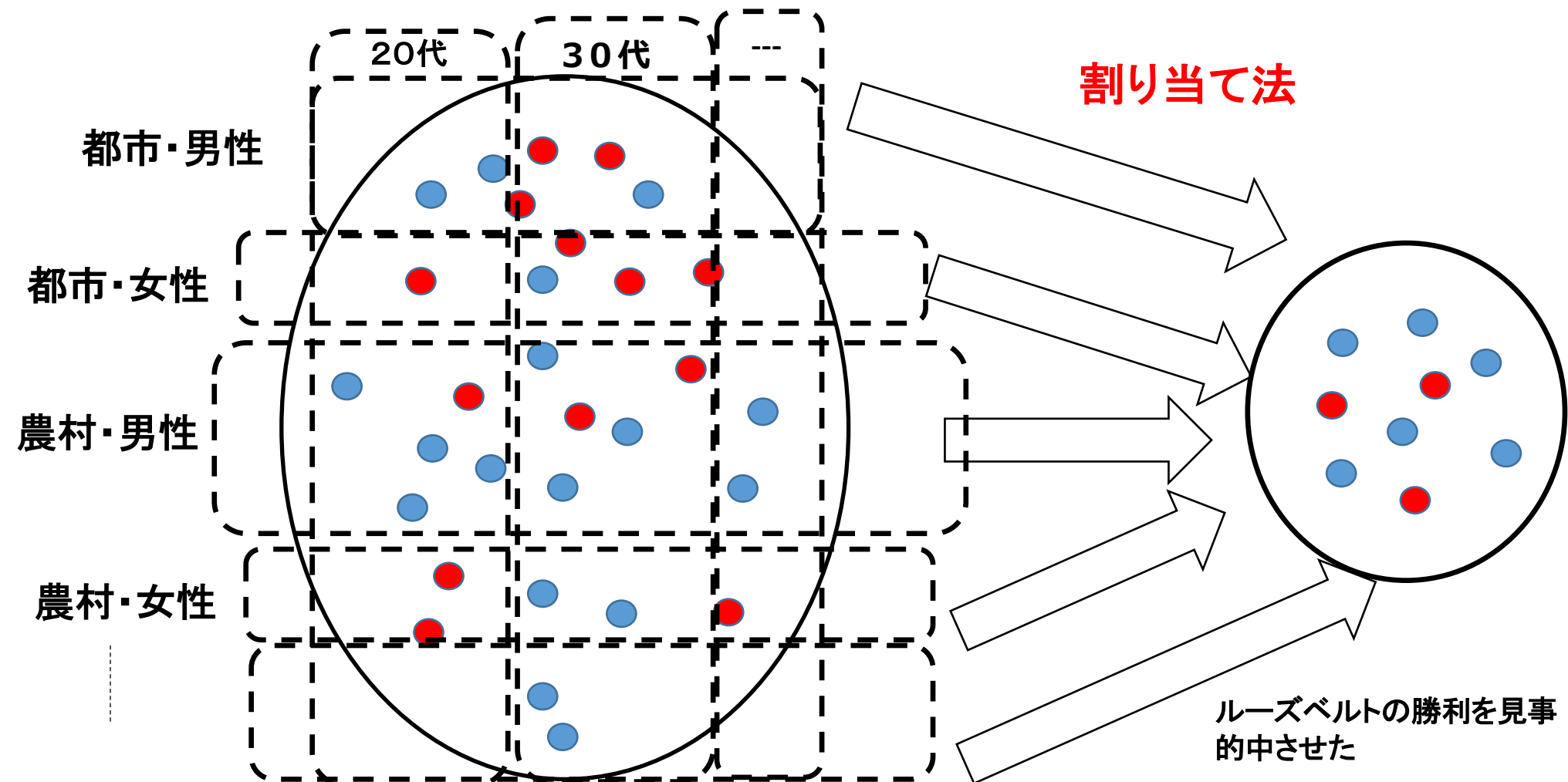
アメリカ世論研究所

3000人を対象に調査を行い、ルーズベルト候補が54%の得票を得て当選することを予想

リテラリー・ダイジェストの抽出方法



アメリカ世論研究所の抽出方法



色々なサンプリング方法

- 単純ランダムサンプリング
- 多段サンプリング
- 層別サンプリング
- 集落サンプリング
- 系統サンプリング

問題

ある大学の学生を対象に、住まいや通学に関するアンケート調査を行う。

- ・ 学生全員に対しての調査は難しい
- ・ 時間や労力をかけないように100人に対して調査



ランダムなサンプリングを行う

単純サンプリング

- 母集団全体から無作為にサンプリングをする方法。
- 人が操作できない偶然によって選び出すようにする。

方法 1



1 0 0 人分の当たりくじ

方法2



10面あるサイコロを投げて、その番号を学生を調査

メリット：人の意思が入ってこないで、ランダム抽出できる

デメリット：対象者全員の参加あるいは事前情報が必要となる

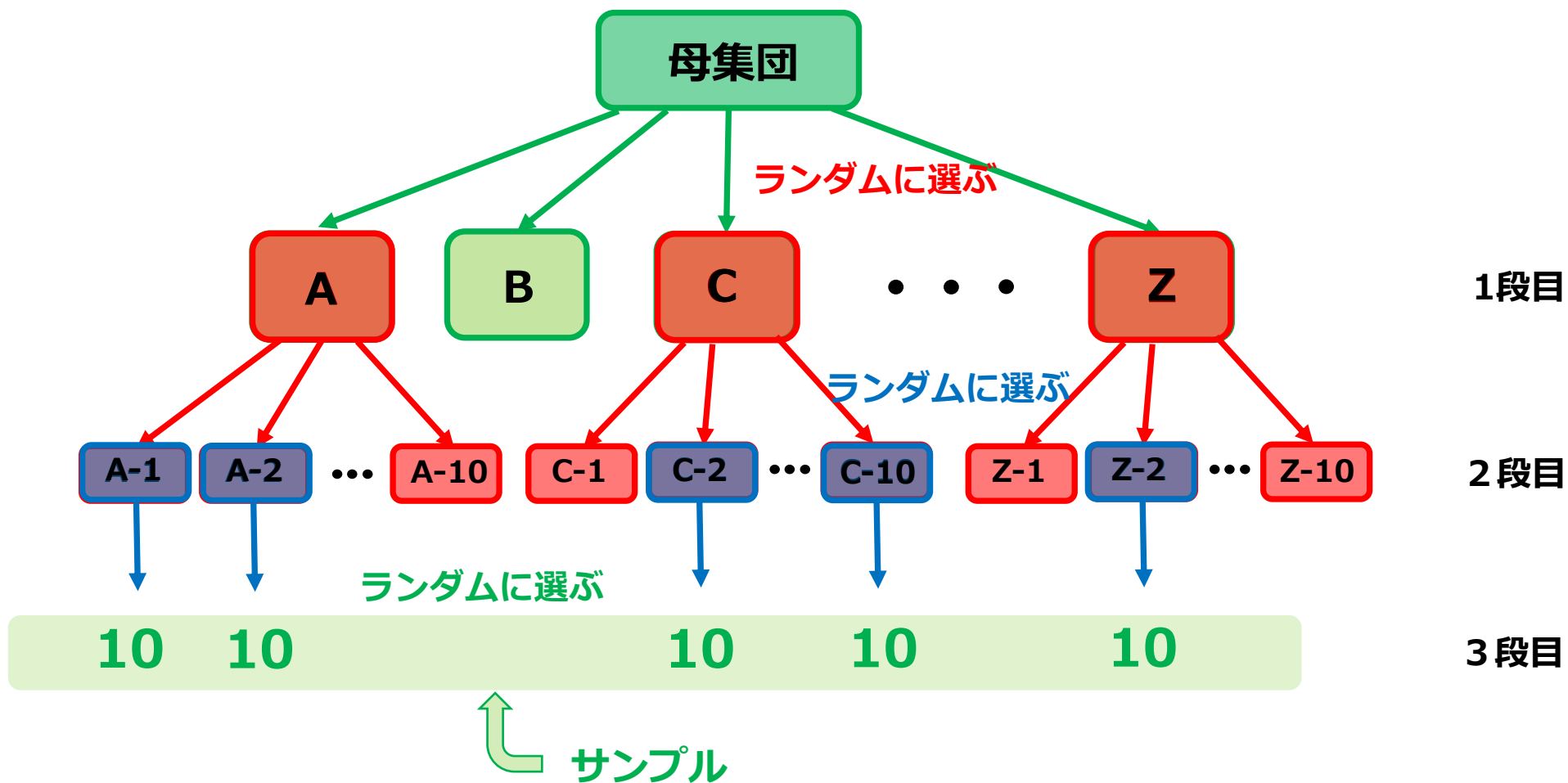
多段サンプリング

- ① 母集団をいくつかのグループに分ける。
- ② いくつかのグループを無作為に選ぶ（1段目）
- ③ さらに細かいグループを無作為に選ぶ（2段目）
- ④ これを繰り返して最終的に無作為に対象を選ぶ。

多段サンプリング

母集団

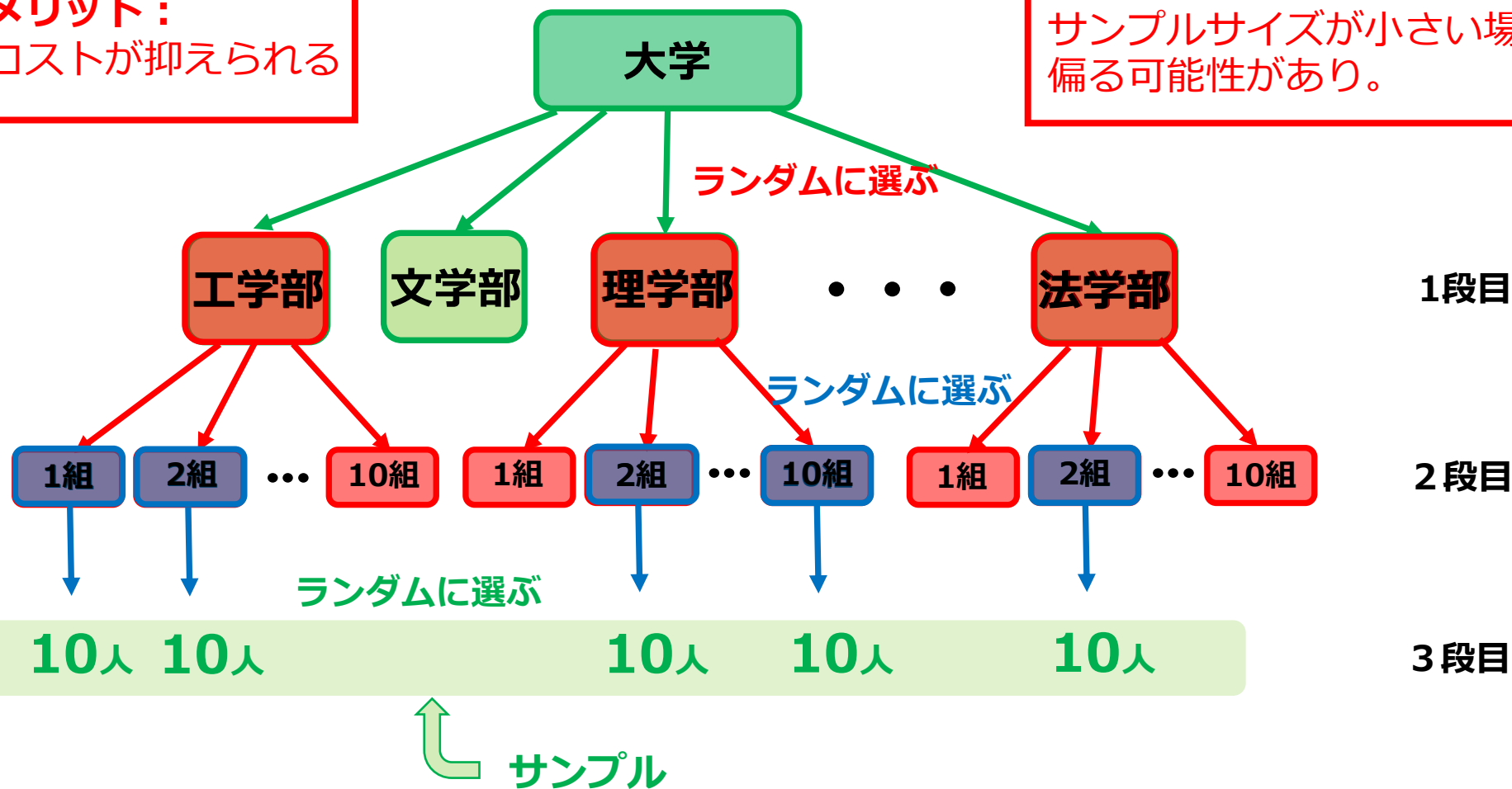
多段サンプリング



多段サンプリング

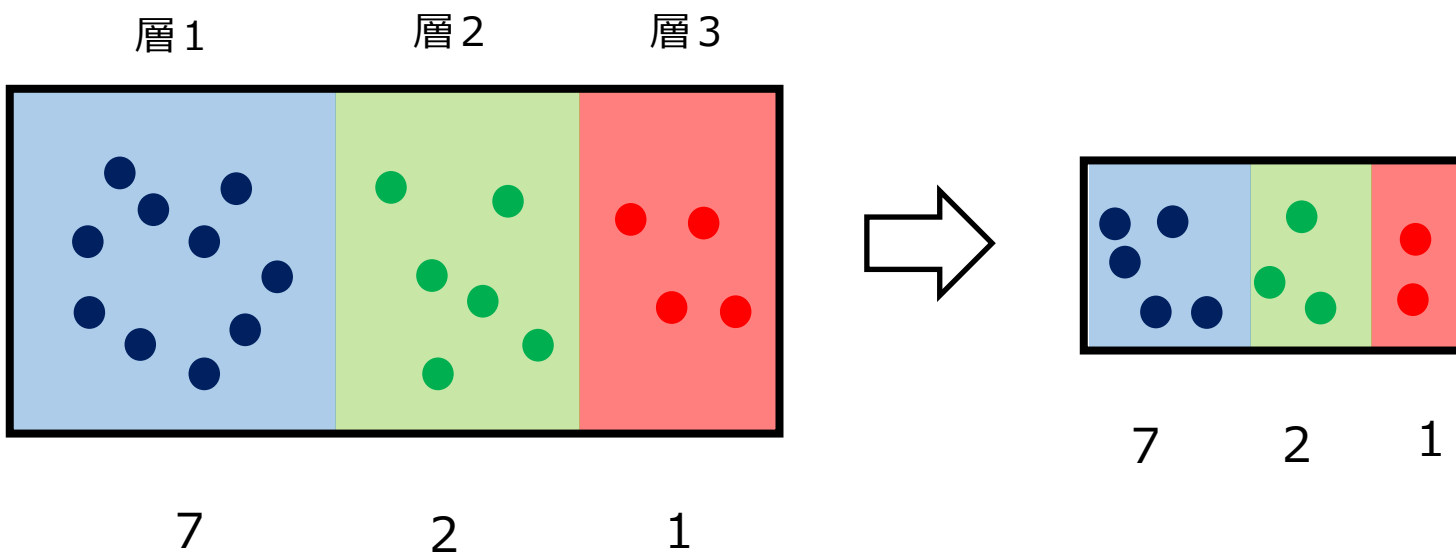
メリット：
コストが抑えられる

デメリット：
サンプルサイズが小さい場合、
偏る可能性がある。



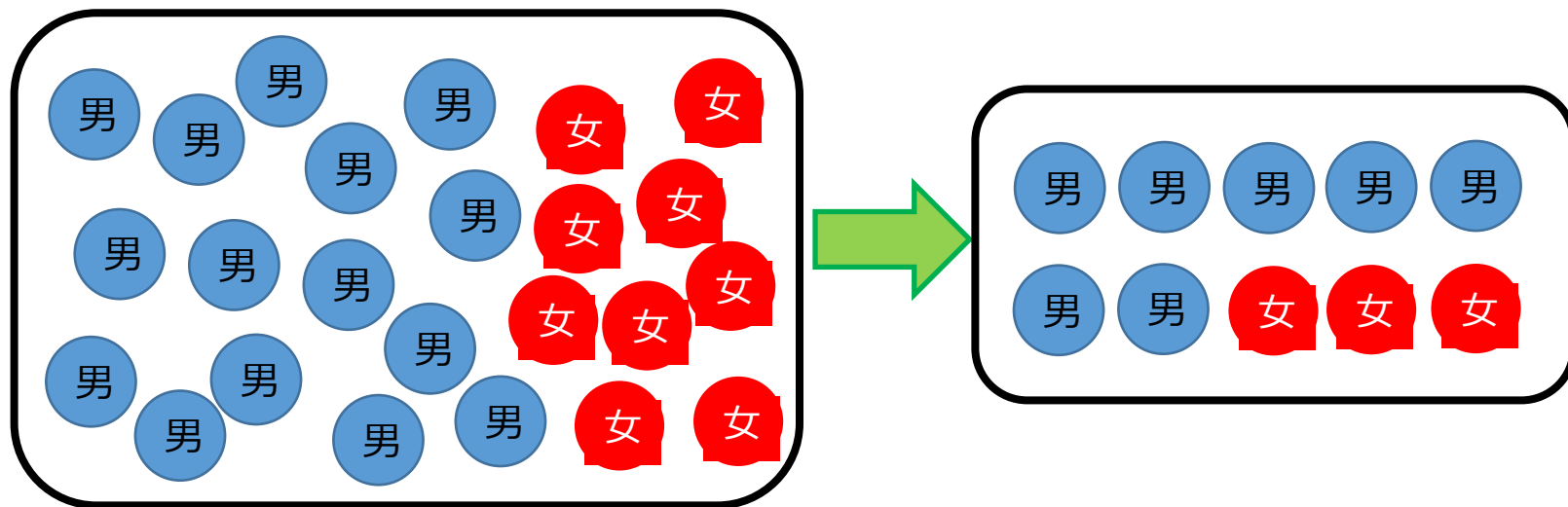
層別サンプリング

母集団をあらかじめ**特徴の異なるいくつかの層**（グループ）に分けておき、各層の中から必要な数の調査対象を無作為に抽出する方法



層別サンプリング

男女比が7:3の大学で、100人の学生を調査する場合、男子70名、女子30名をそれぞれに無作為に抽出する。



層別サンプリング

メリット：

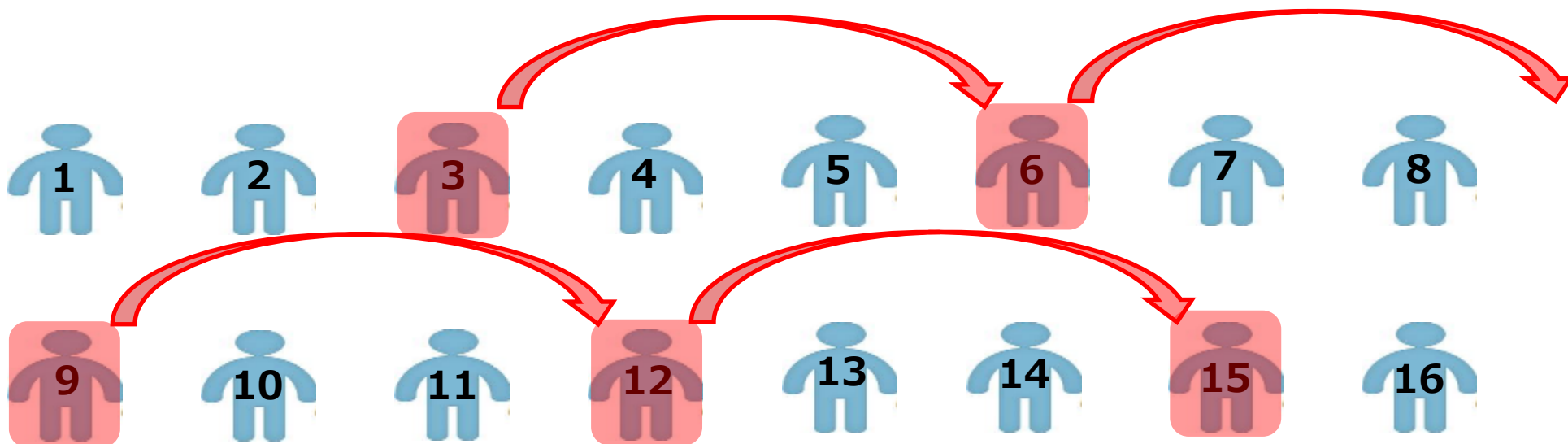
- ・ 母集団内情報（年齢別、性別など）の比較を行える
- ・ 母集団の推測の精度が増す
- ・ 各層の特徴が大きく異なる場合に有用

デメリット：

- ・ 母集団の構成情報を事前に知っておく必要がある

系統サンプリング

- ① 母集団のデータに通し番号を付ける。
- ② サンプルを一つ無作為に選び、そこから一定の間隔をあけて番号を選んでいき、サンプルを抽出していく。

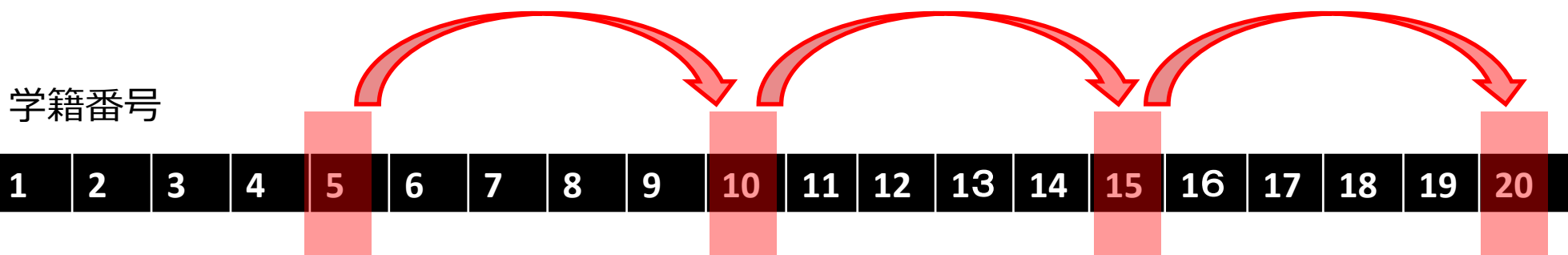


【例】視聴率調査

系統サンプリング

最初だけ無作為に抽出し、間隔を決めておけば、動的にサンプリングしてくれる。

【例】大学生の調査では学籍番号が利用できる。



メリット：

単純無作為抽出より手間や時間やコストが掛からない。最初の1つだけ選べばOK

デメリット：

名簿の並び順に何らかの周期があると標本に偏りが生じる可能性がある。

クラスターサンプリング（集落サンプリング）

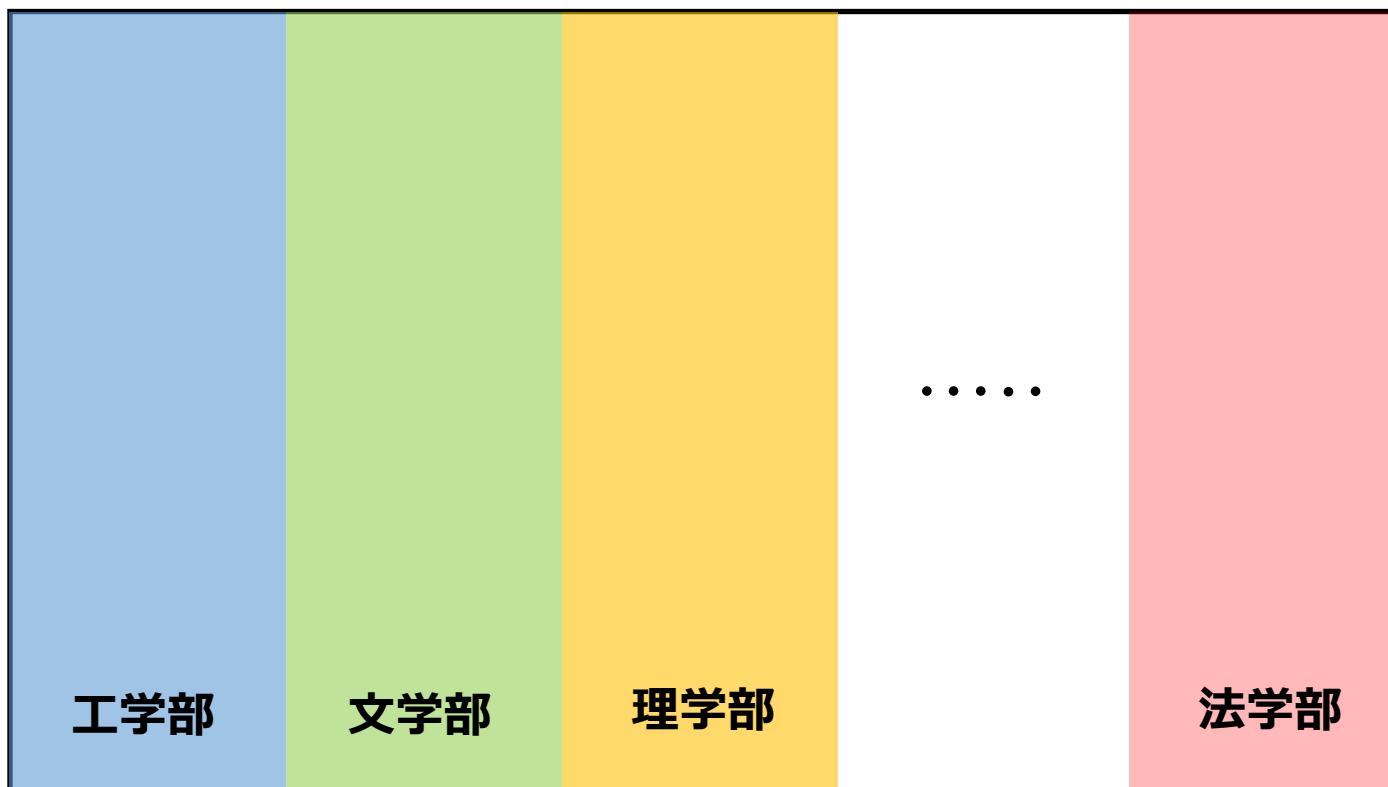
- ① 母集団を何らかの基準でグループ分けする。
- ② グループをランダムに選び、選ばれたグループの要素を**すべて**調べる方法。

グループ間の特徴の違いは小さく、
グループ内の特徴の違いは大きくなるような設定をする

クラスターサンプリング（集落サンプリング）

【例】

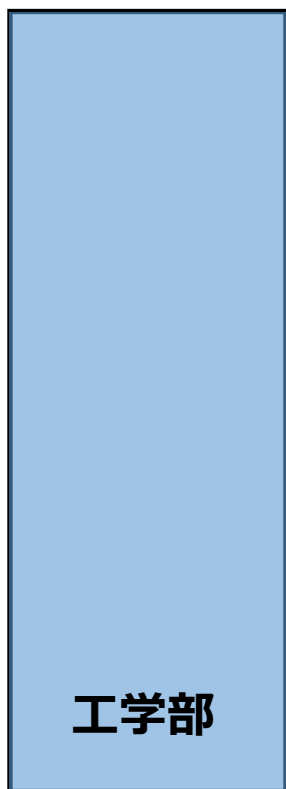
大学



クラスターサンプリング（集落サンプリング）

【例】

大学



工学部



理学部

全調査を行う

データ分析の視点

- データの持つ情報を数値やグラフを使って表現する

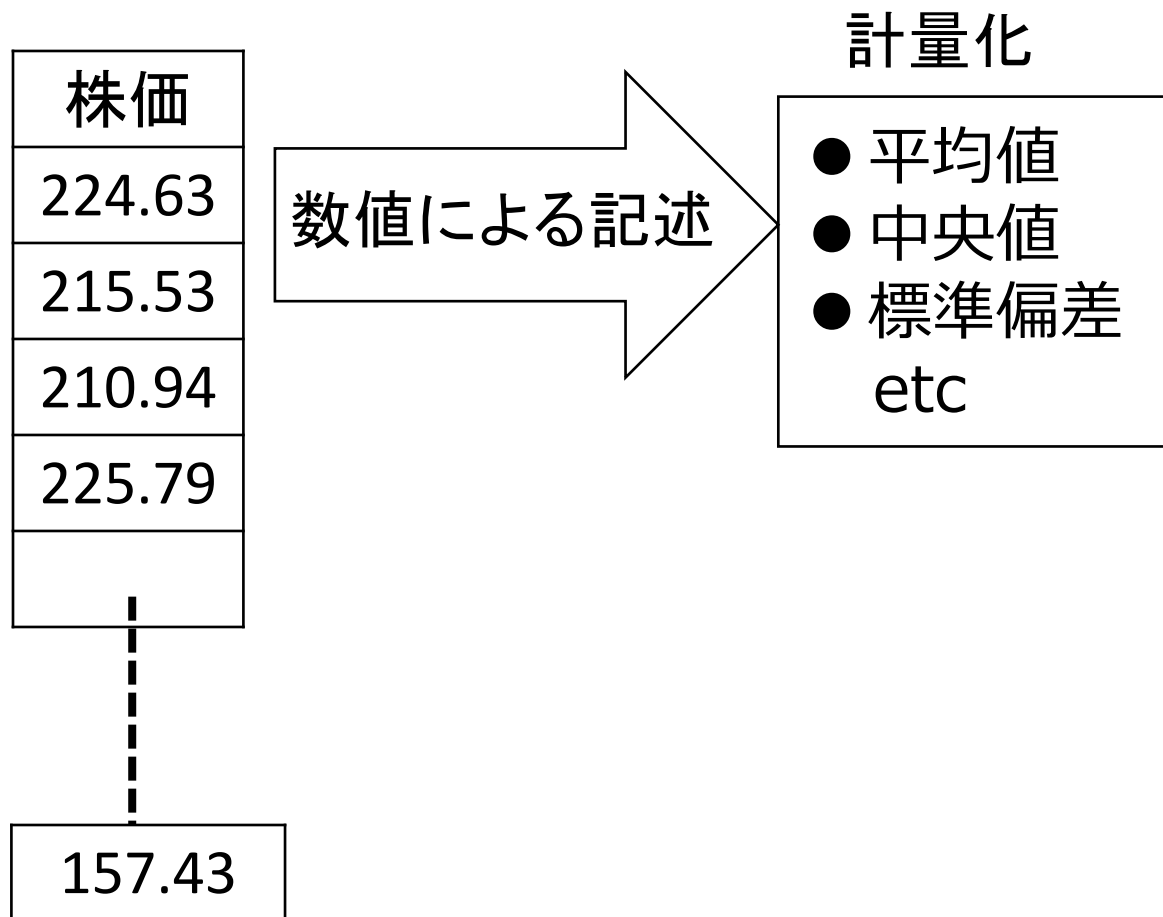
データ分析の視点

- データの持つ情報を**数値**や**グラフ**を使って表現する

株価
224.63
215.53
210.94
225.79
157.43

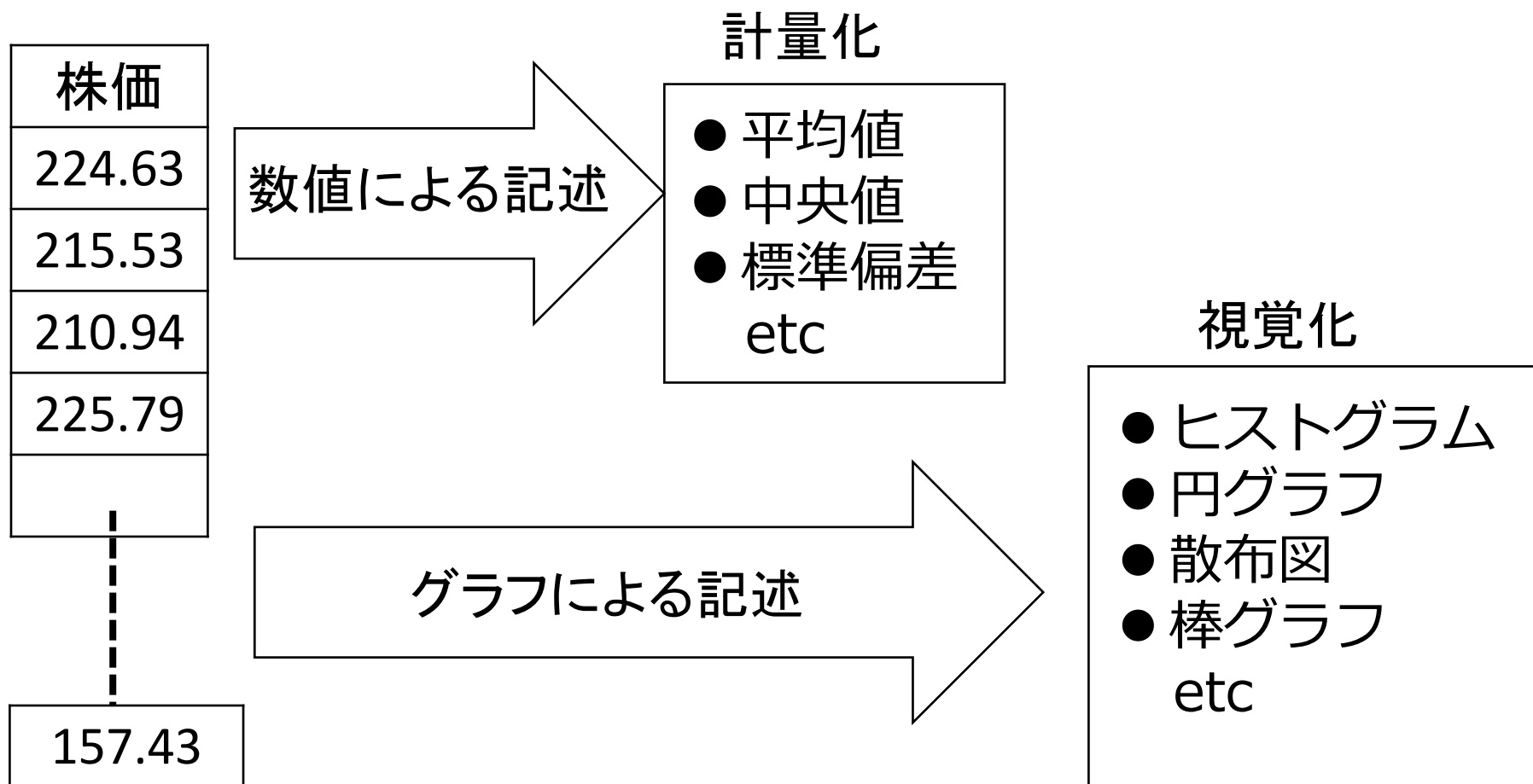
データ分析の視点

- データの持つ情報を**数値**や**グラフ**を使って表現する



データ分析の視点

- データの持つ情報を**数値**や**グラフ**を使って表現する



度数分布表（質的データの場合）

データ値



成績	度数	相対度数
A		
B		
C		
D		

度数分布表（質的データの場合）

データ値



観測値の数



成績	度数	相対度数
A	10	
B	16	
C	10	
D	4	

合計 40

度数分布表（質的データの場合）

データ値



観測値の数



構成比率

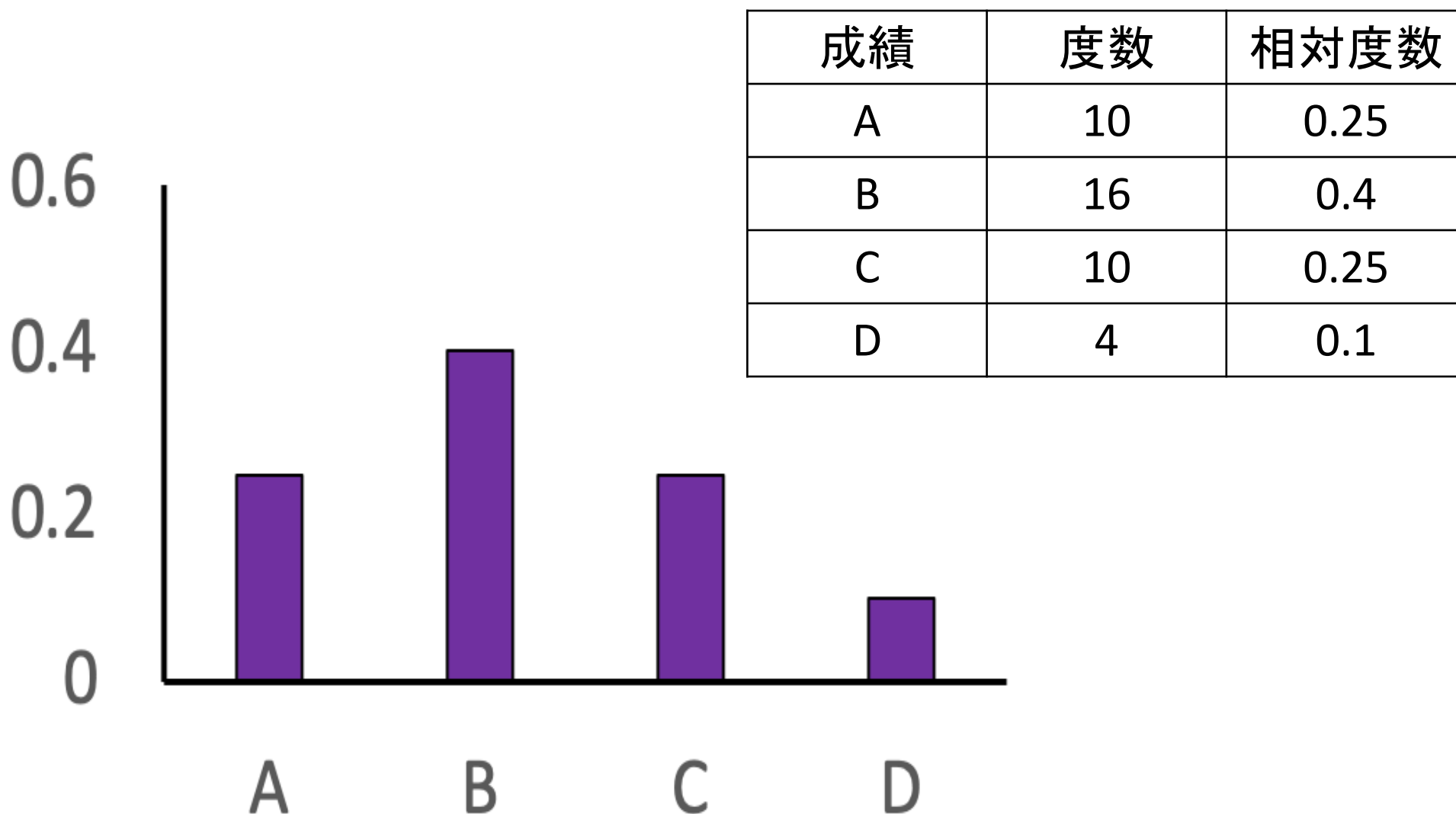


成績	度数	相対度数
A	10	0.25
B	16	0.4
C	10	0.25
D	4	0.1

合計 40

合計 1

ヒストグラム（質的データの場合）



度数分布表（量的データの場合）

株価
24.5
15.3
10.9
5.79
98.4

度数分布表（量的データの場合）

株価
24.5
15.3
10.9
5.79

48.4

データ値



区間	度数	相対度数
10-20		
20-30		
30-40		
40-50		

度数分布表（量的データの場合）

株価
24.5
15.3
10.9
5.79

48.4

データ値



観測値の数



区間	度数	
10-20	10	
20-30	16	
30-40	10	
40-50	4	

合計 40

度数分布表（量的データの場合）

株価
24.5
15.3
10.9
5.79

48.4

データ値



観測値の数



構成比率

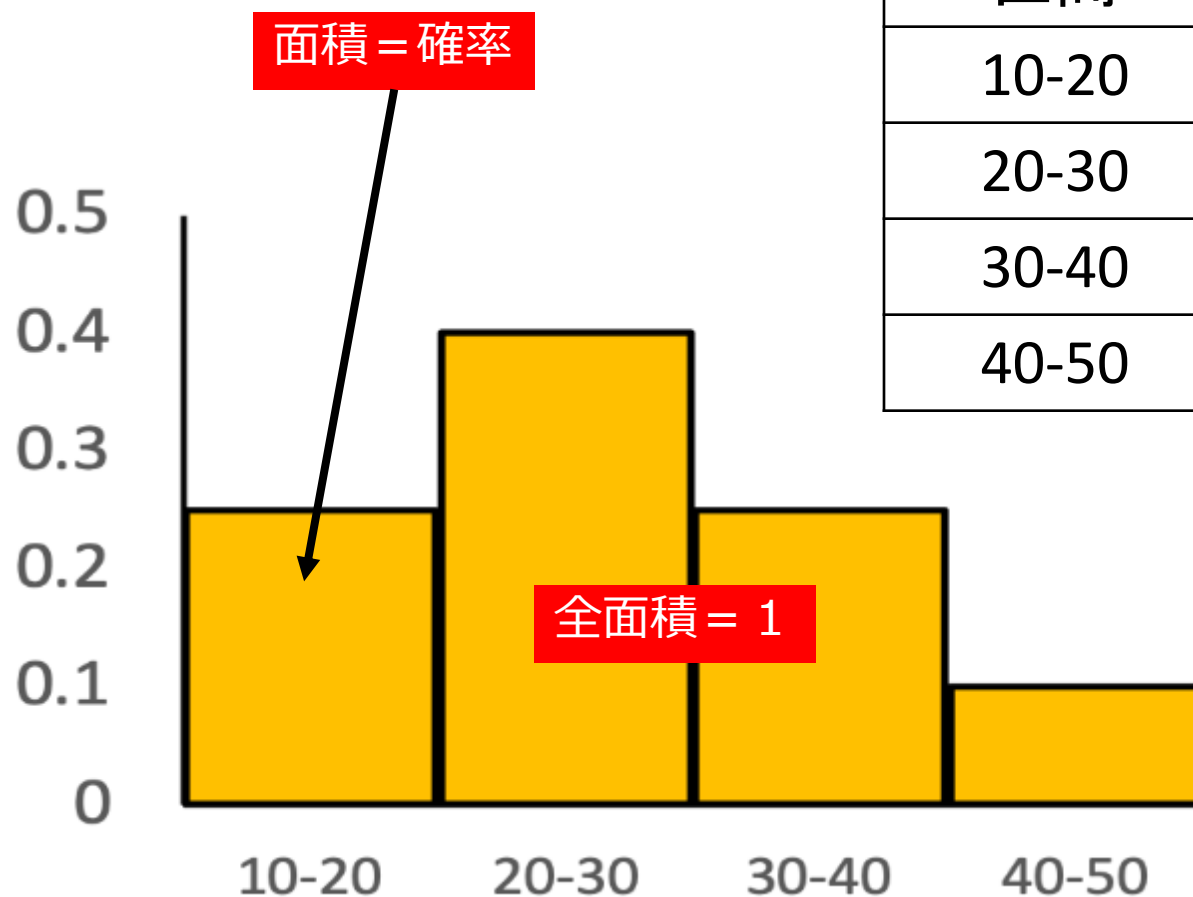


区間	度数	相対度数
10-20	10	0.25
20-30	16	0.4
30-40	10	0.25
40-50	4	0.1

合計 40

合計 1

ヒストグラム（量的データの場合）

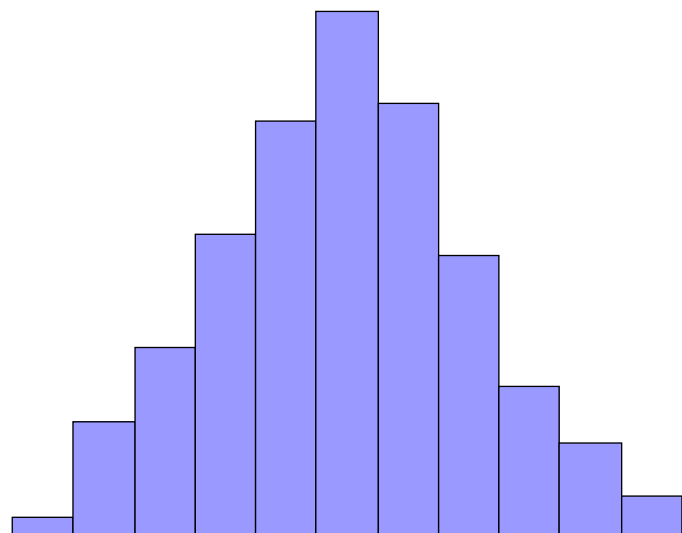


ヒストグラムの意味

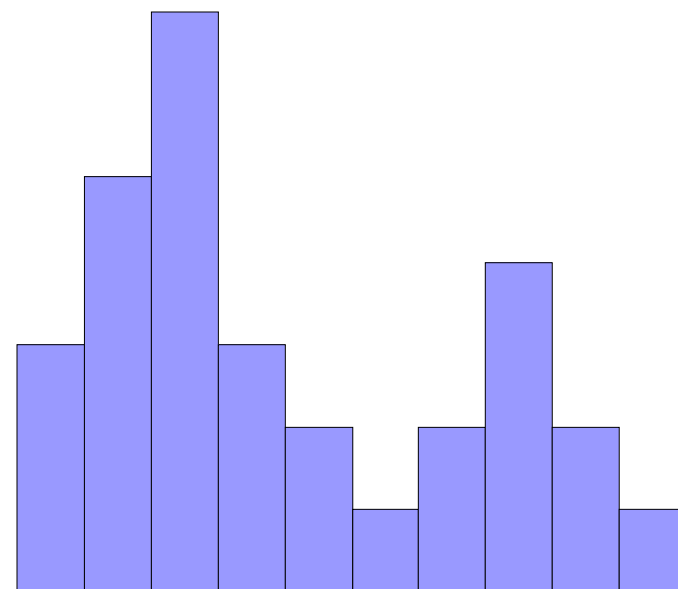
- 3つの構成要素
 - 横軸：データの数値軸
 - 棒の高さ：起りやすさ（注目すべき事柄）
 - 棒の面積：割合・度数
- 区間幅が同じであれば、高さと面積が対応
- **幅が異なる場合、注意が必要**
 - （練習問題参照）

分布のチェックポイント

単峰 か 多峰 か？



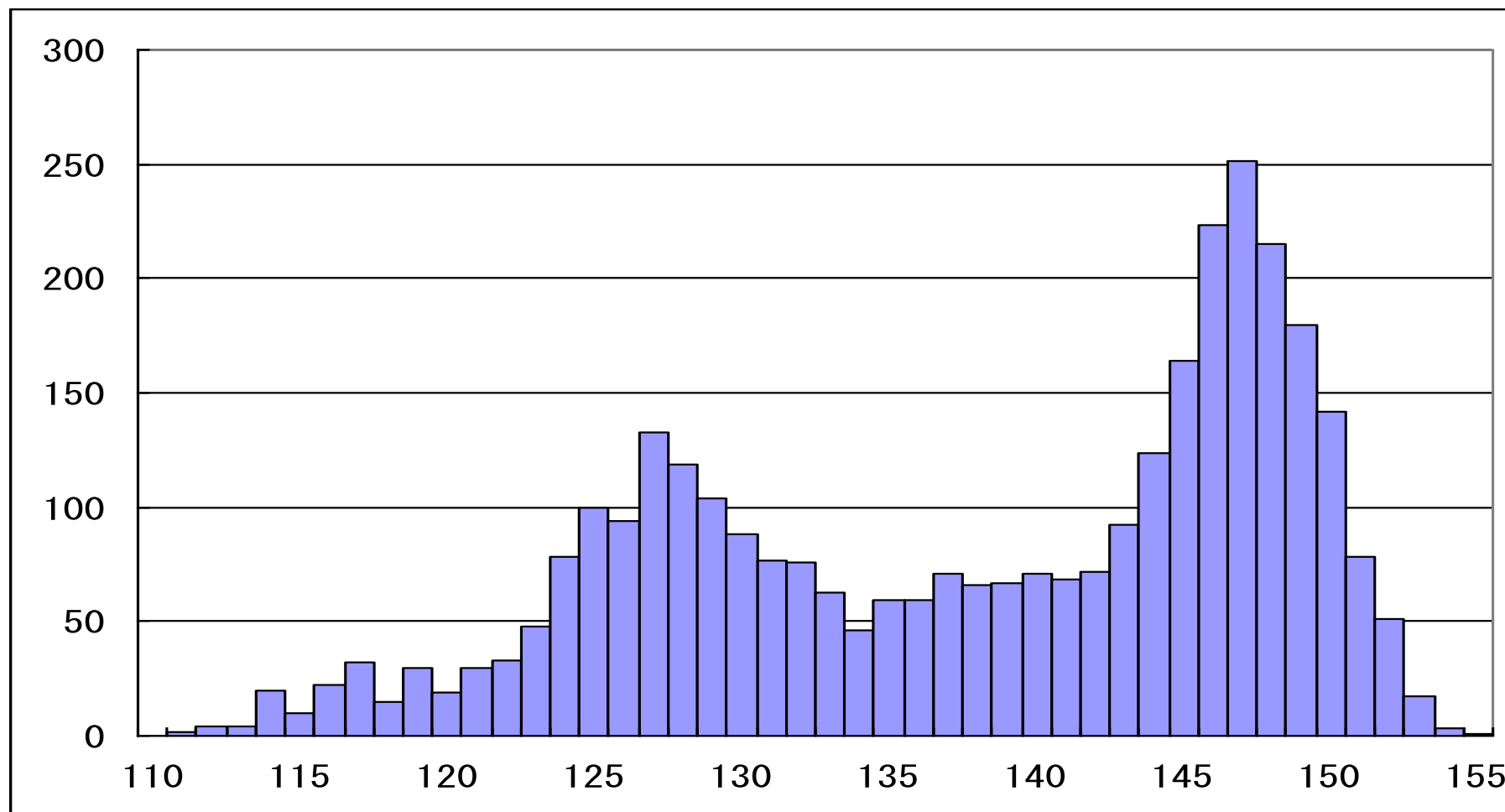
単峰で左右対称なヒストグラム



2峰性のヒストグラム

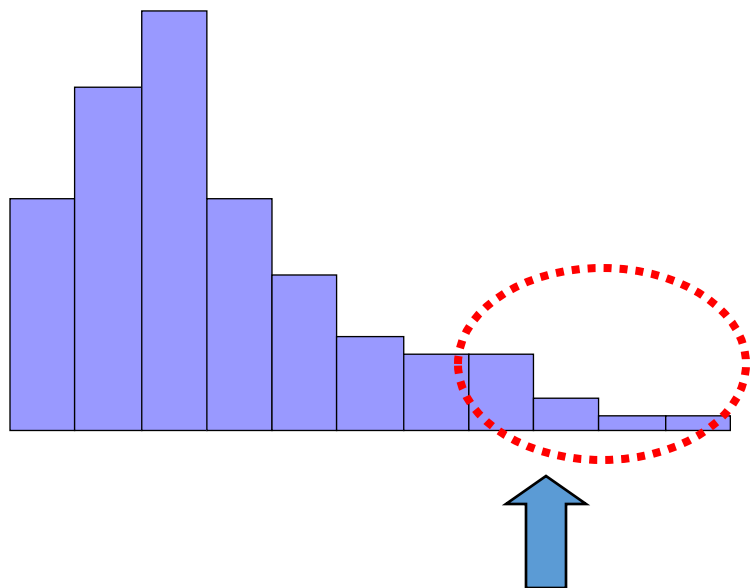
2 峰性の例

- 松坂投手の球速のヒストグラム



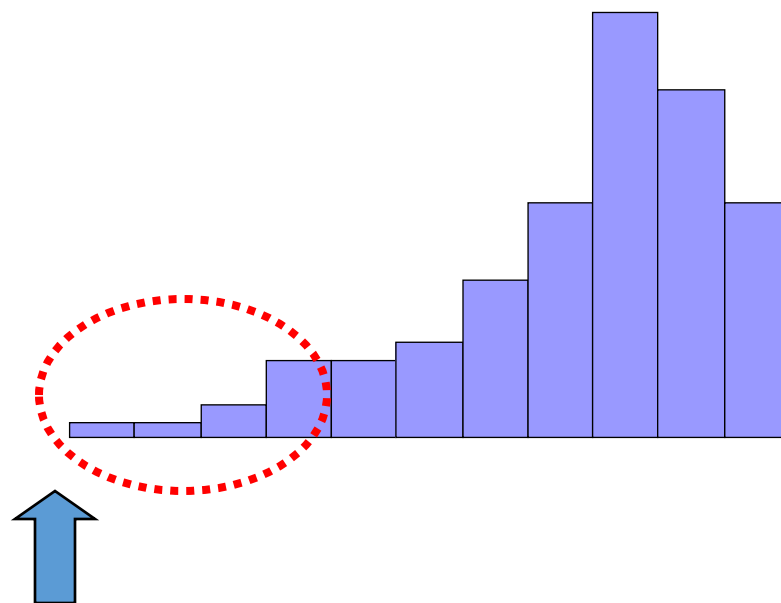
分布のチェックポイント

- ・ 対称 か 非対称 か



右方向に裾をひく分布

右に歪んだ分布



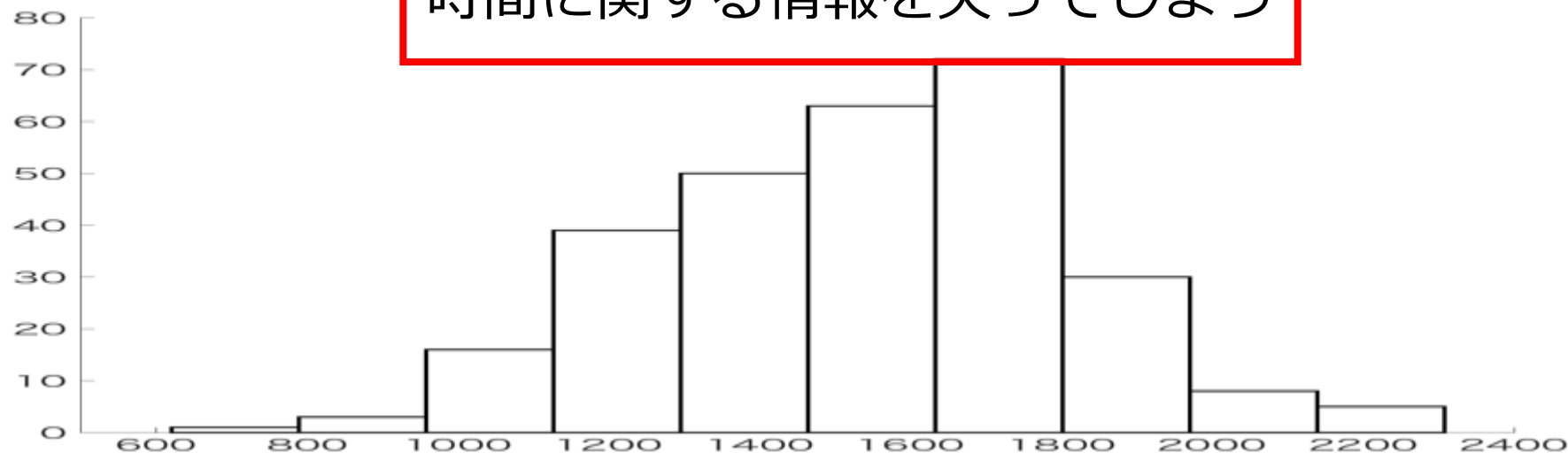
左方向に裾をひく分布

左に歪んだ分布

ヒストグラムを使った分析例



時間に関する情報を失ってしまう



クラスターに分解する

日付	応募者数
6月1日	657
6月2日	814
6月3日	718
6月4日	554
6月5日	569
6月6日	664
6月7日	1059
6月8日	950

クラスターに分解する

日付	応募者数	曜日
6月1日	657	木曜
6月2日	814	金曜
6月3日	718	土曜
6月4日	554	日曜
6月5日	569	月曜
6月6日	664	火曜
6月7日	1059	水曜
6月8日	950	木曜

クラスターに分解する

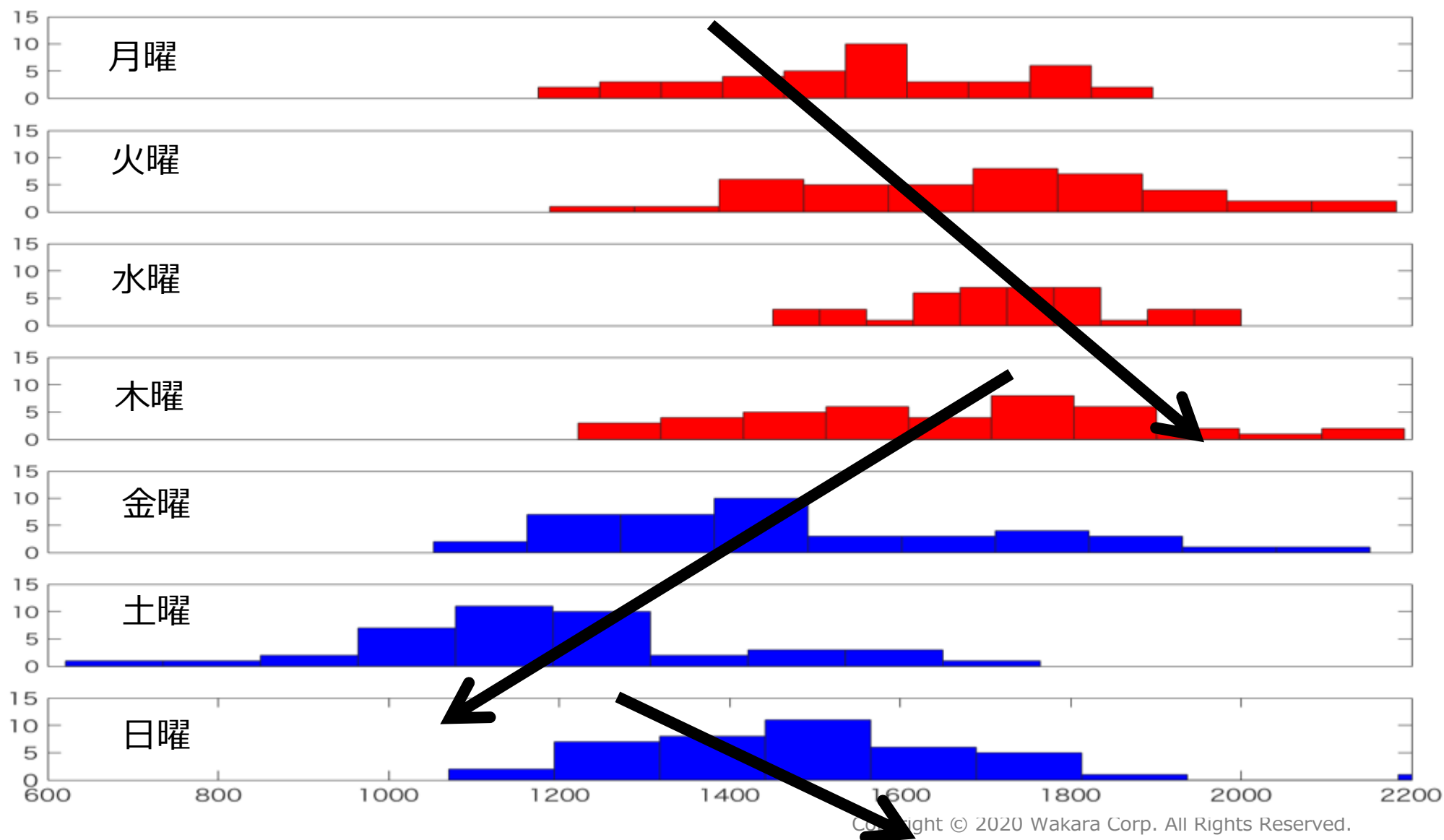
日付	応募者数	曜日
6月1日	657	木曜
6月2日	814	金曜
6月3日	718	土曜
6月4日	554	日曜
6月5日	569	月曜
6月6日	664	火曜
6月7日	1059	水曜
6月8日	950	木曜

曜日毎のデータに分解



木曜日
657
950
1120
1202
982
1239
1220
1056
921

クラスターに分解してヒストグラム



- ここまでは集計だけの分析
- 確率の概念を応用すると、何が出来るのか？

確率論

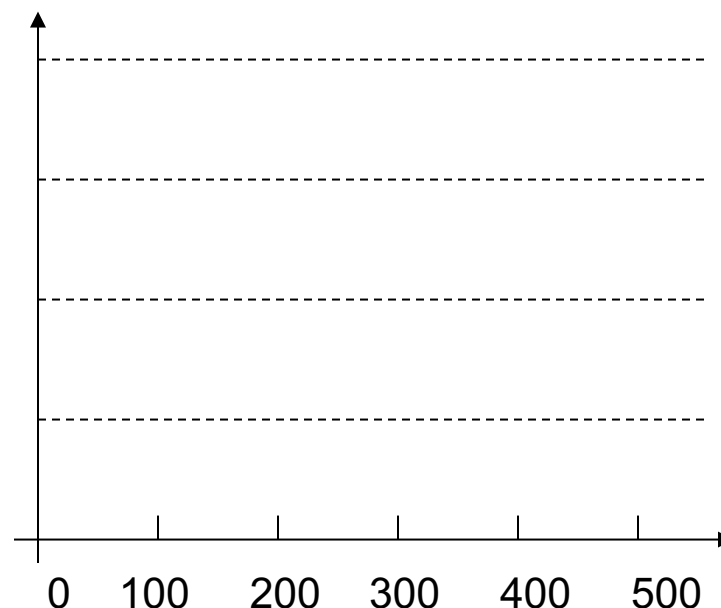
確率の公理

- $0 \leq P(A) \leq 1$
- $P(S) = 1, \text{ or } P(A^c) = 1 - P(A)$
- $A_1, A_2 \cdots A_n$ が排反事象なら $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

問題演習 1

- 次のような集計表が得られた。ヒストグラムを作成してください。

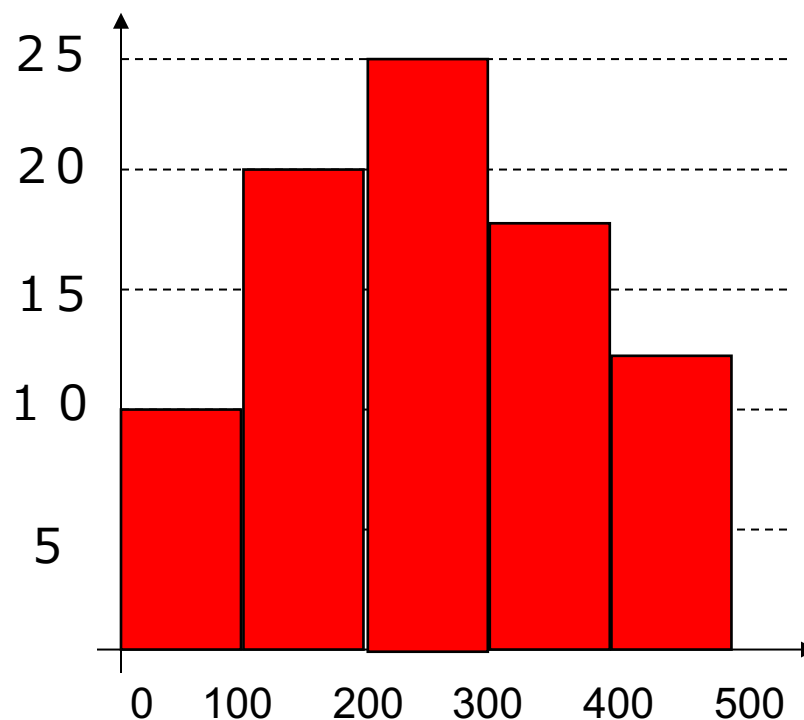
範囲	度数
0-100	10
100-200	20
200-300	25
300-400	18
400-500	12



問題演習 1

- 次のような集計表が得られた。ヒストグラムを作成してください。

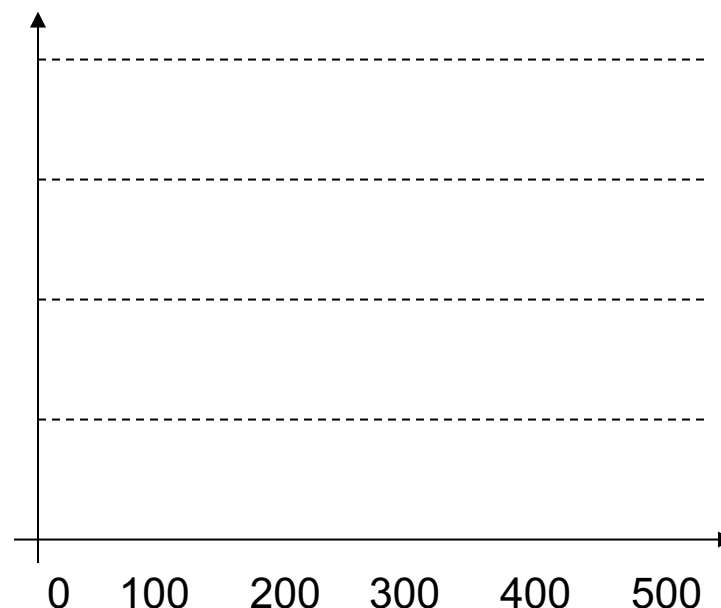
範囲	度数
0-100	10
100-200	20
200-300	25
300-400	18
400-500	12



問題演習 2

- 次のような集計表が得られた。ヒストグラムを作成してください。

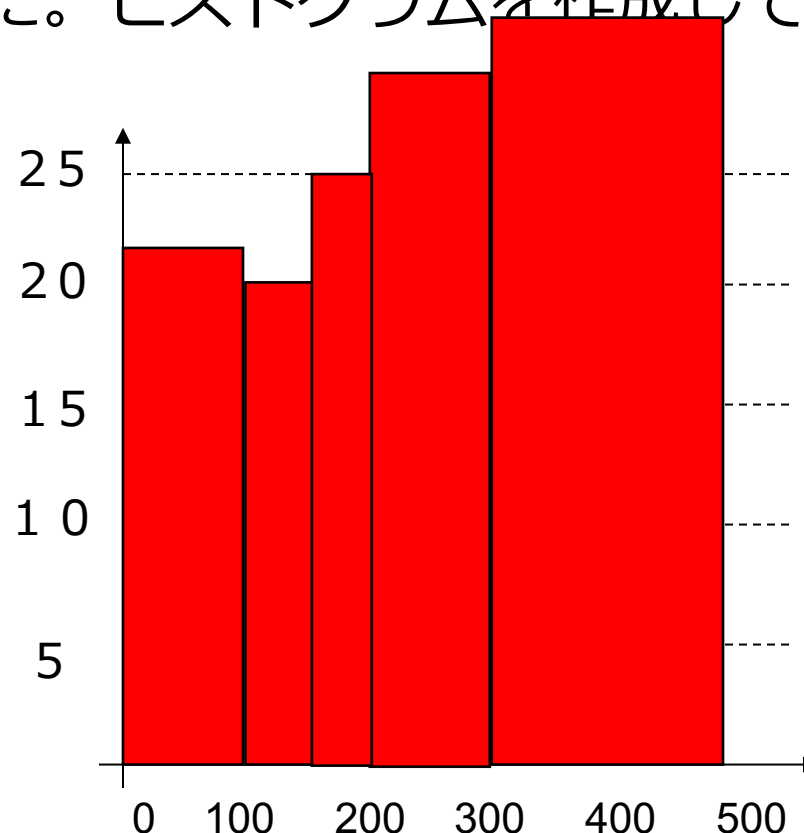
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



問題演習2

- 次のような集計表が得られた。ヒストグラムを作成してください。

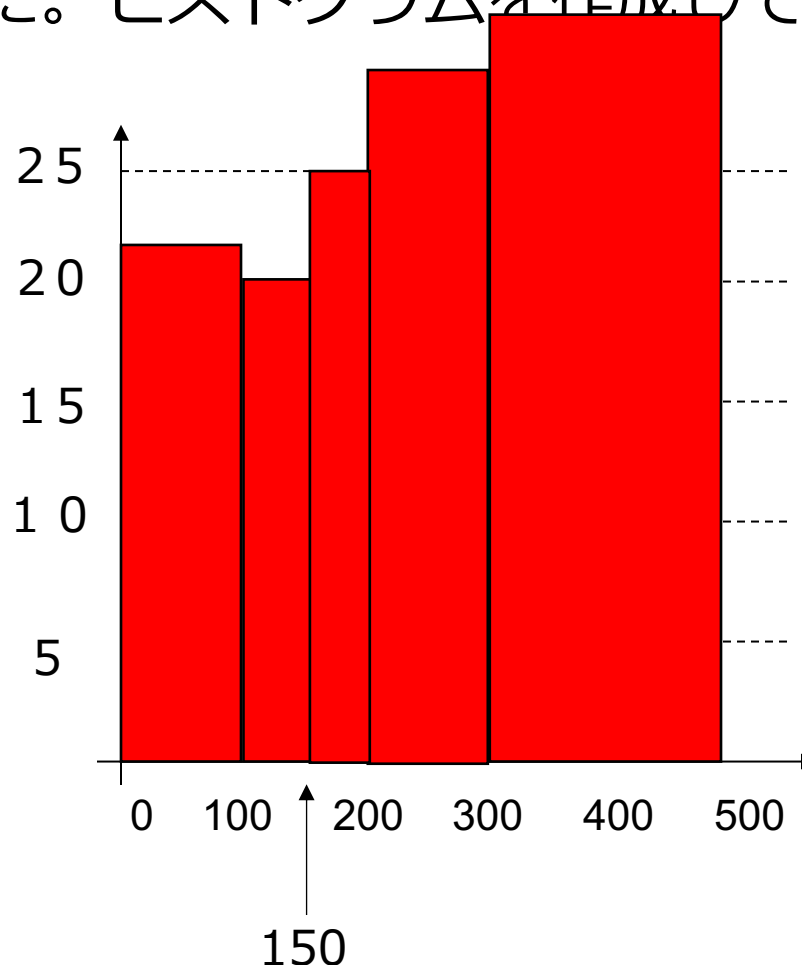
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



問題演習2

- 次のような集計表が得られた。ヒストグラムを作成してください。

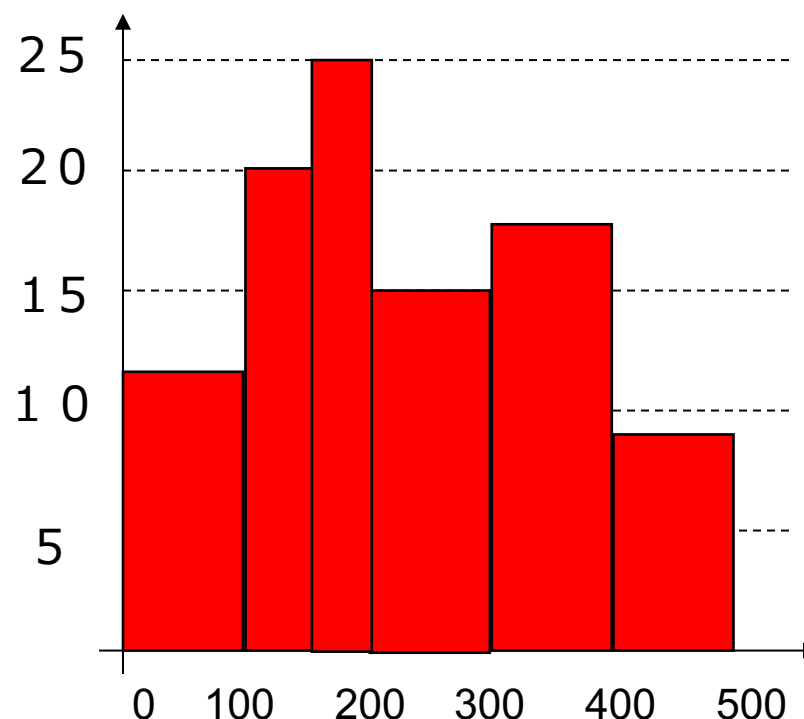
範囲	度数	相対度数
0-100	22	0.17
100-150	20	0.15
150-200	25	0.19
200-300	30	0.23
300-500	32	0.24



問題演習2

- 次のような集計表が得られた。ヒストグラムを作成してください。

範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



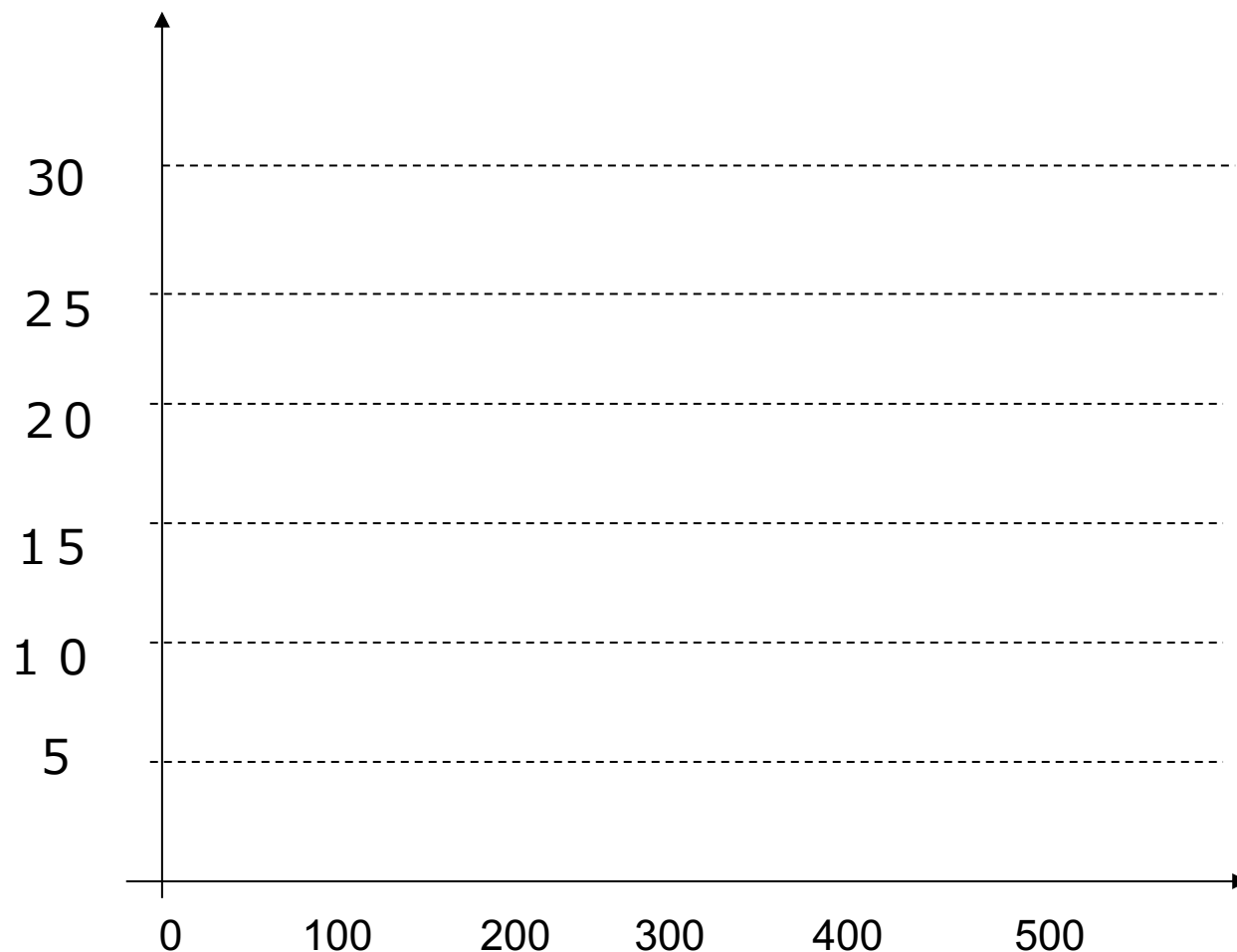
問題演習2

範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32

範囲が一番小さいのは？

幅 = 50 \longleftrightarrow 幅 = 1

(この幅を基準にする)

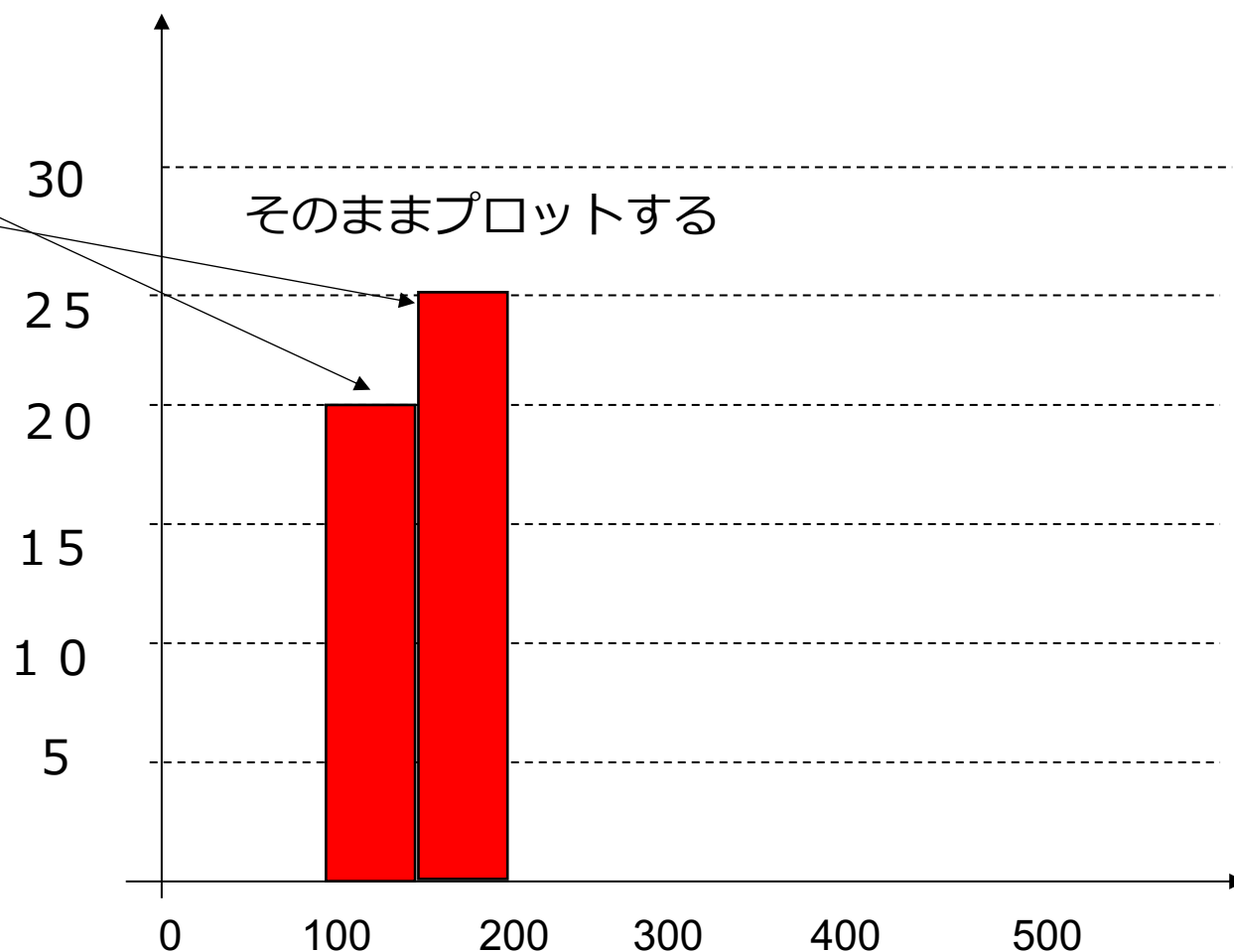


問題演習2

範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32

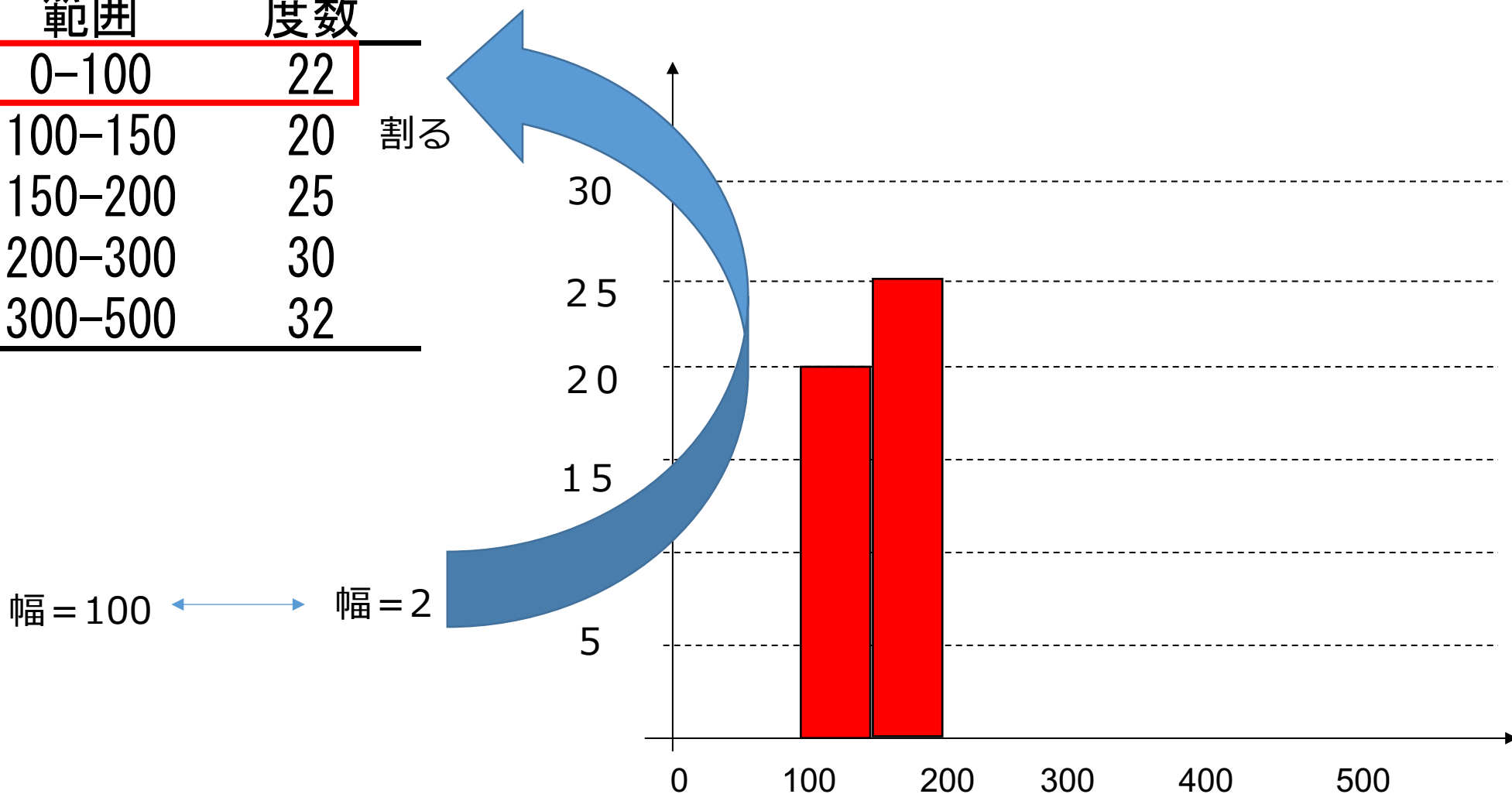
範囲が一番小さいのは？

幅 = 50



問題演習2

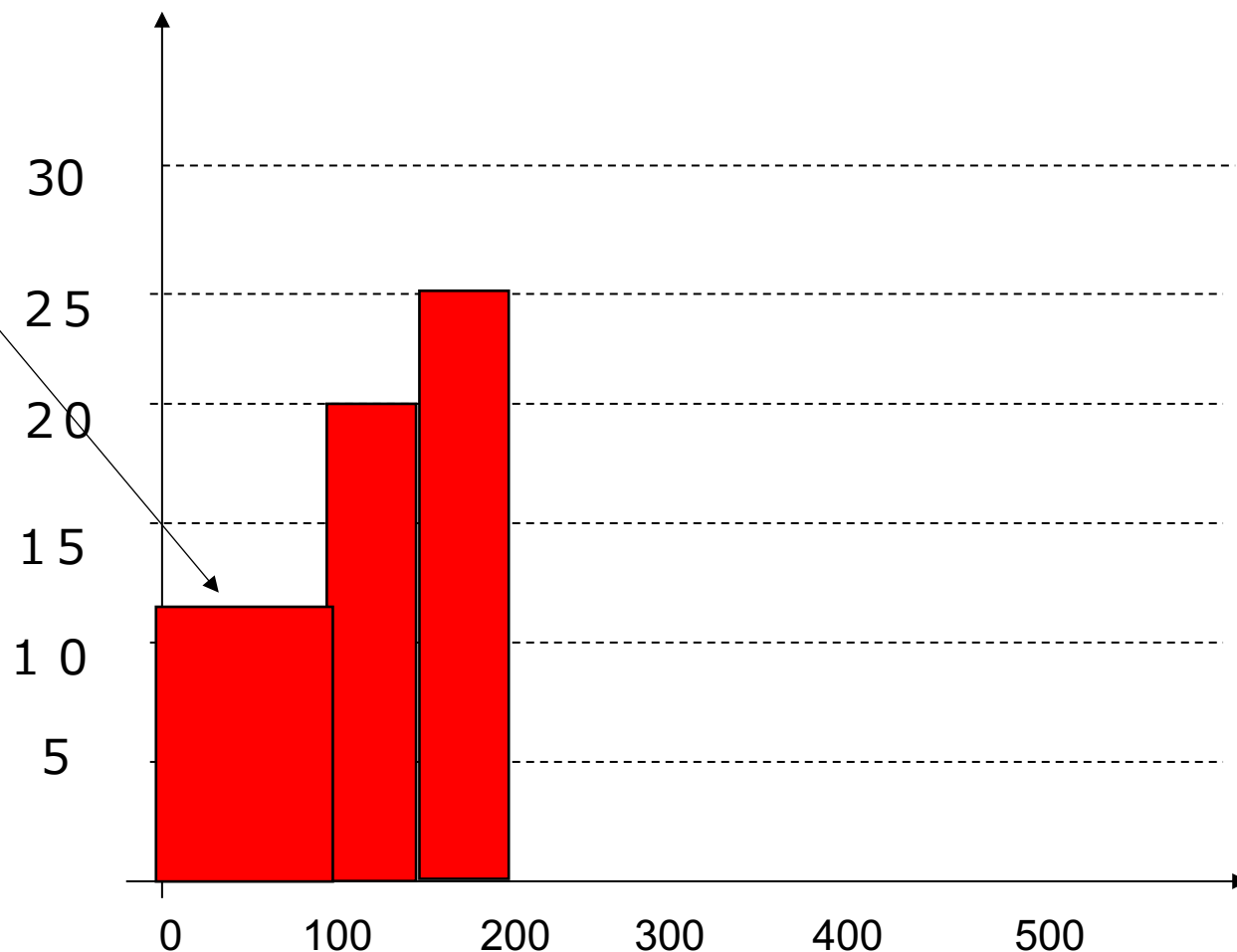
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



問題演習2

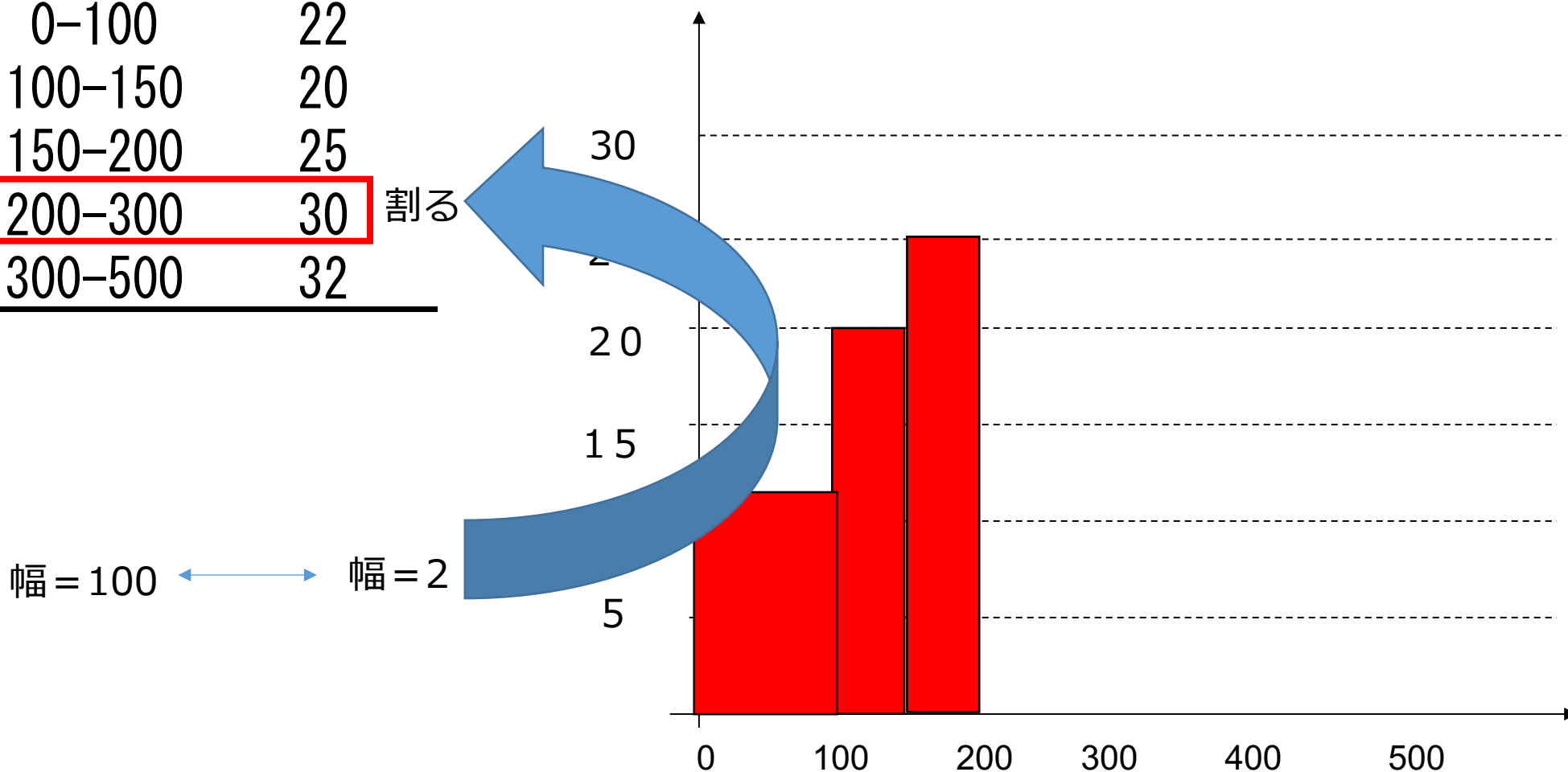
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32

割った数をプロットする



問題演習2

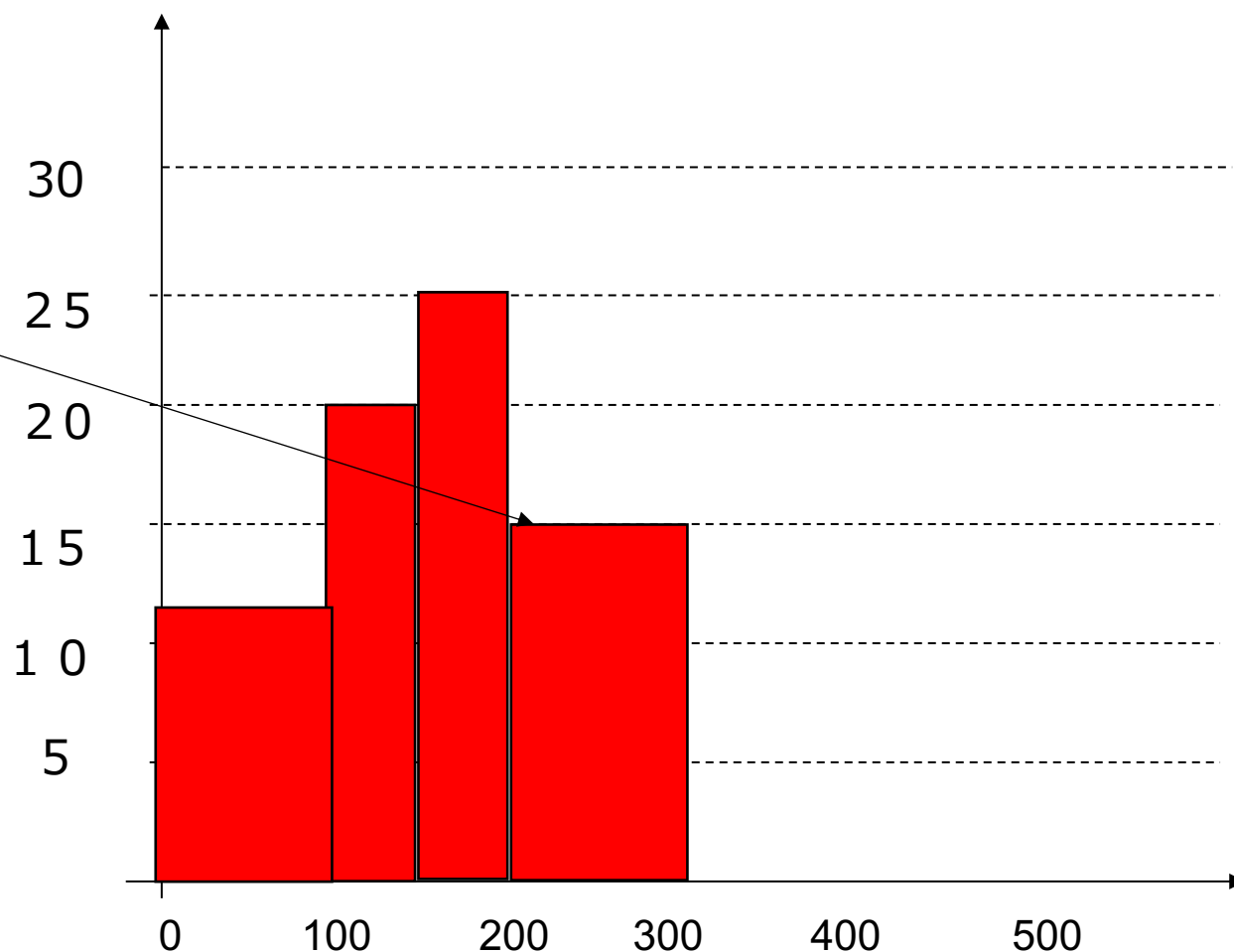
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



問題演習2

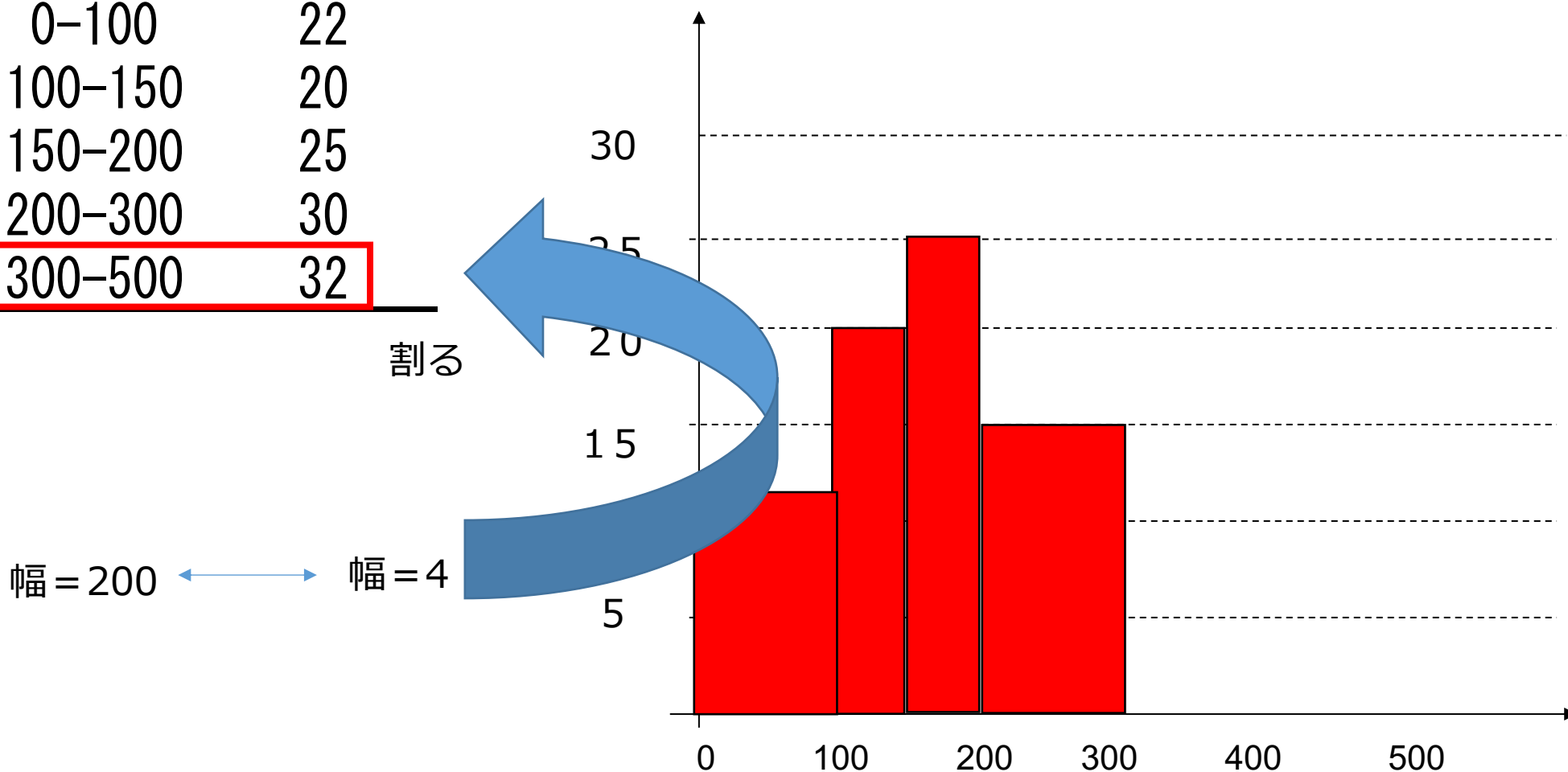
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32

割った数をプロットする



問題演習2

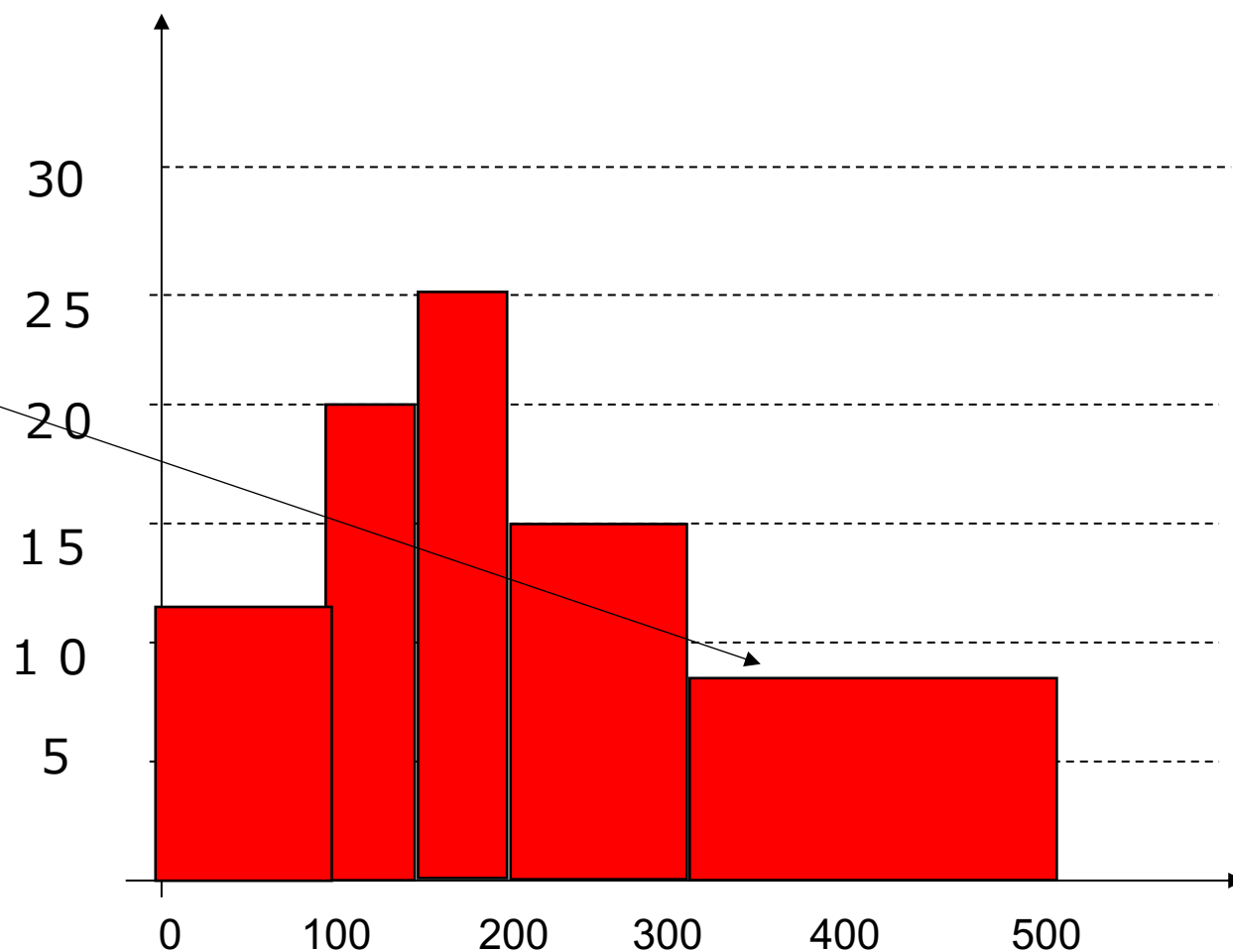
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



問題演習2

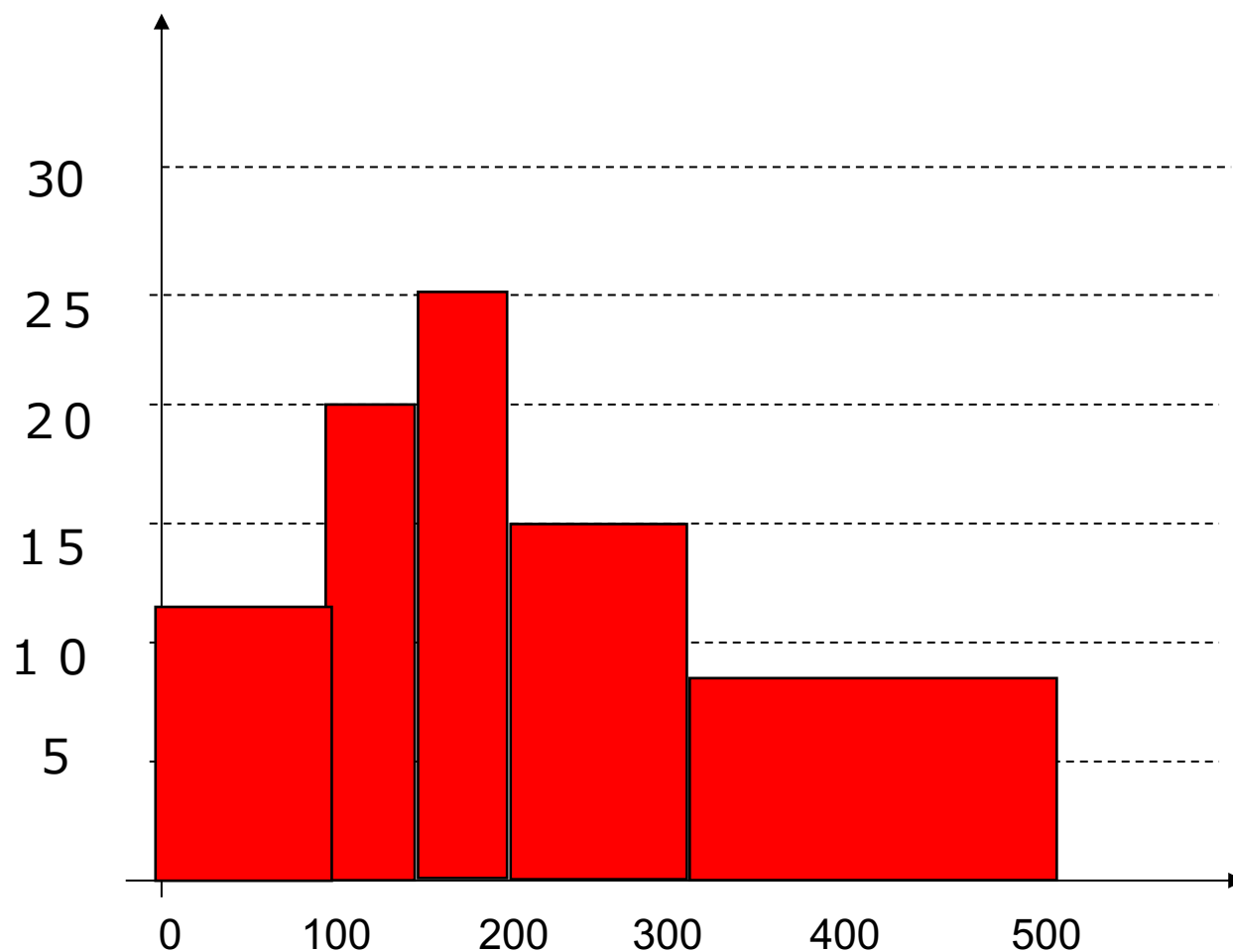
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32

割った数をプロットする



問題演習2

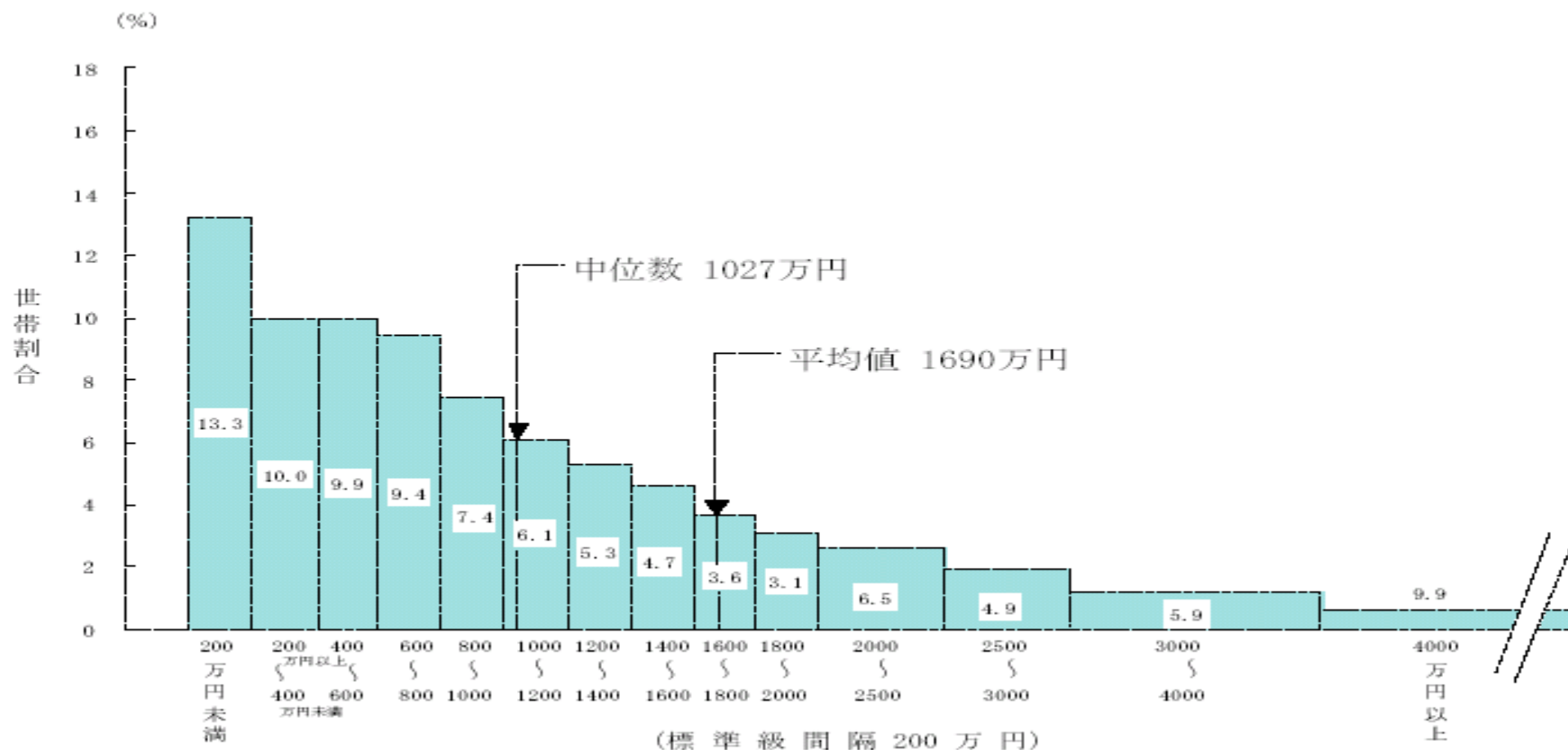
範囲	度数
0-100	22
100-150	20
150-200	25
200-300	30
300-500	32



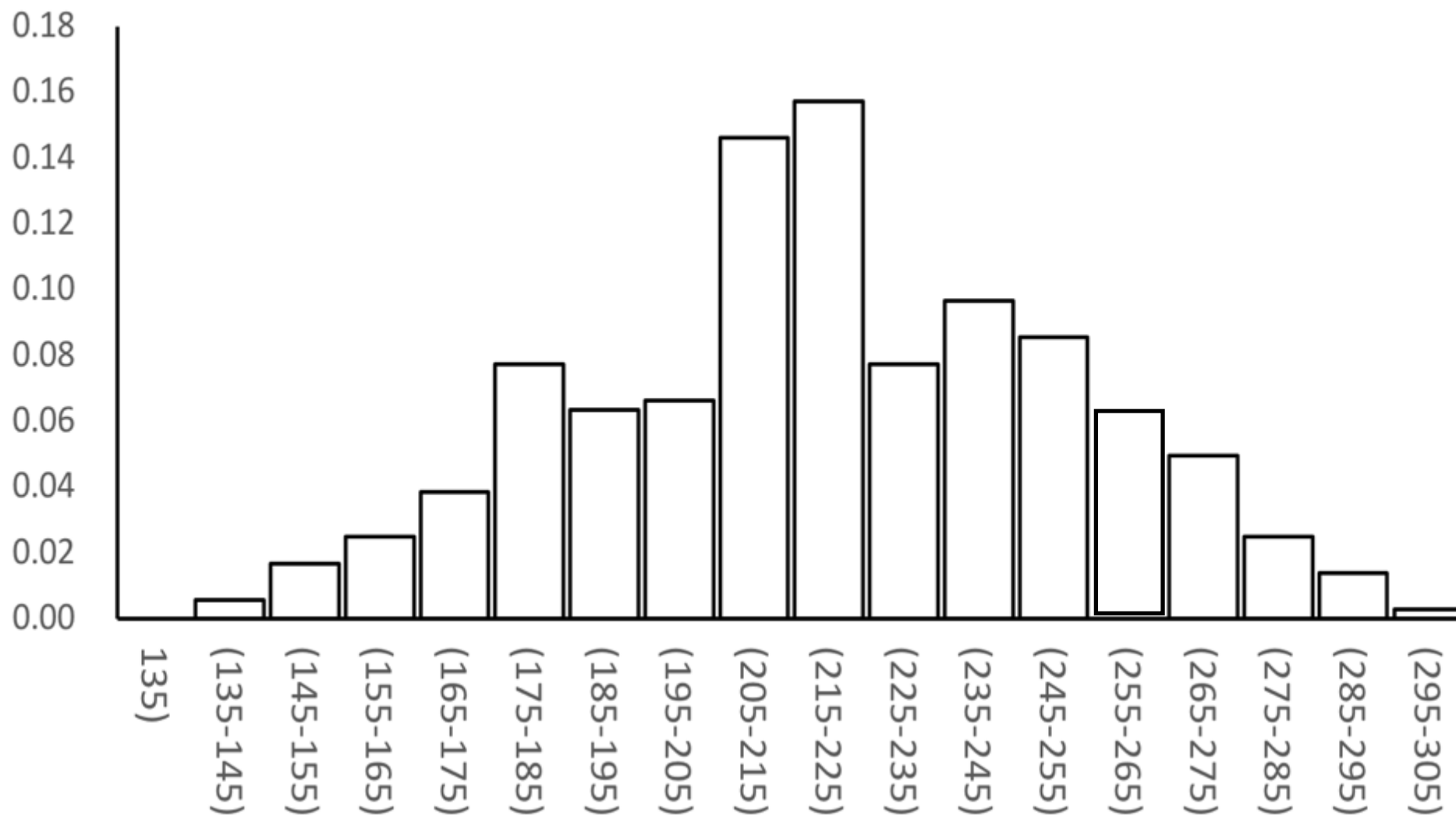
歪んだ分布の例

- 世帯貯蓄金額のヒストグラム

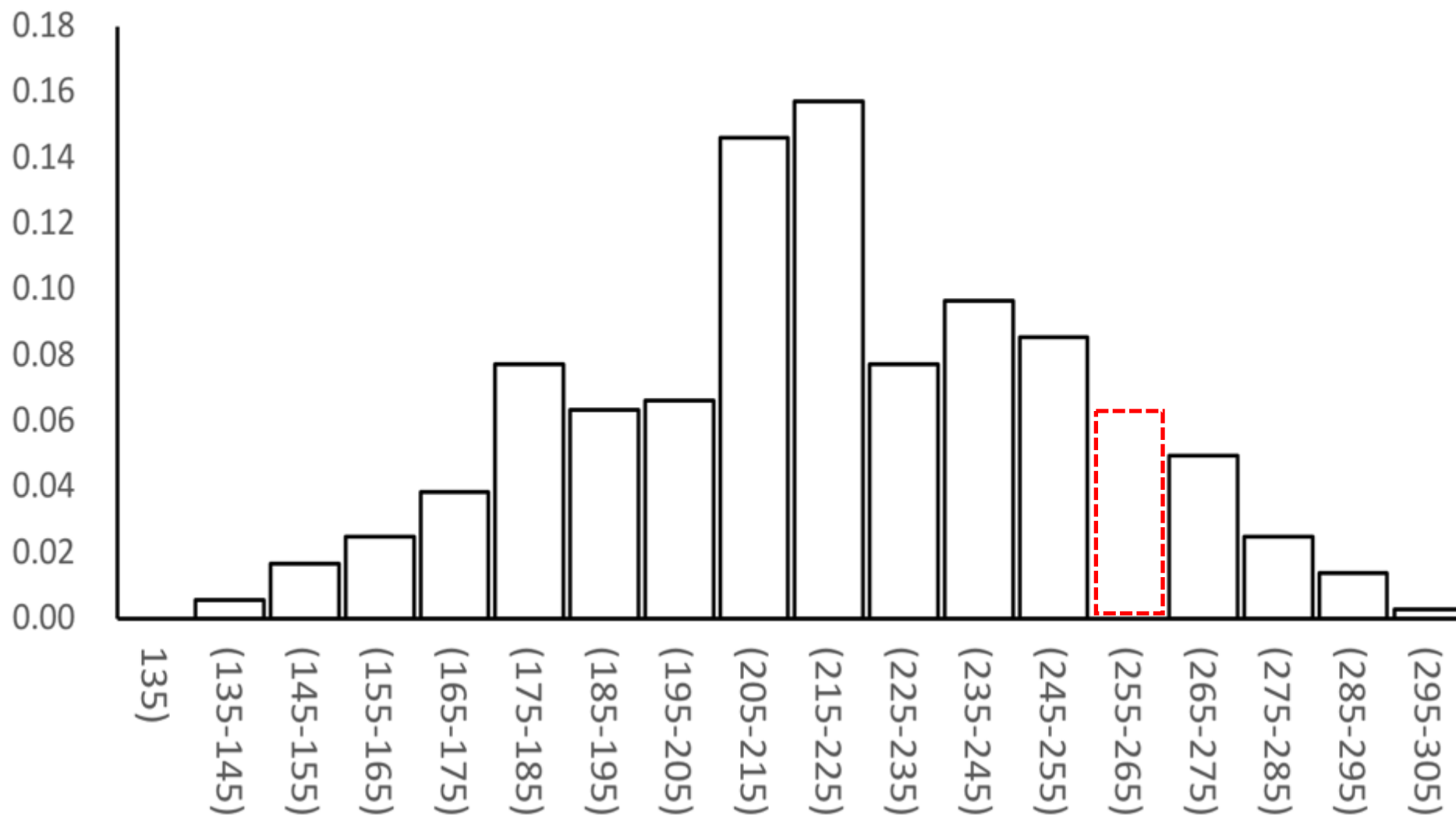
(全世帯)



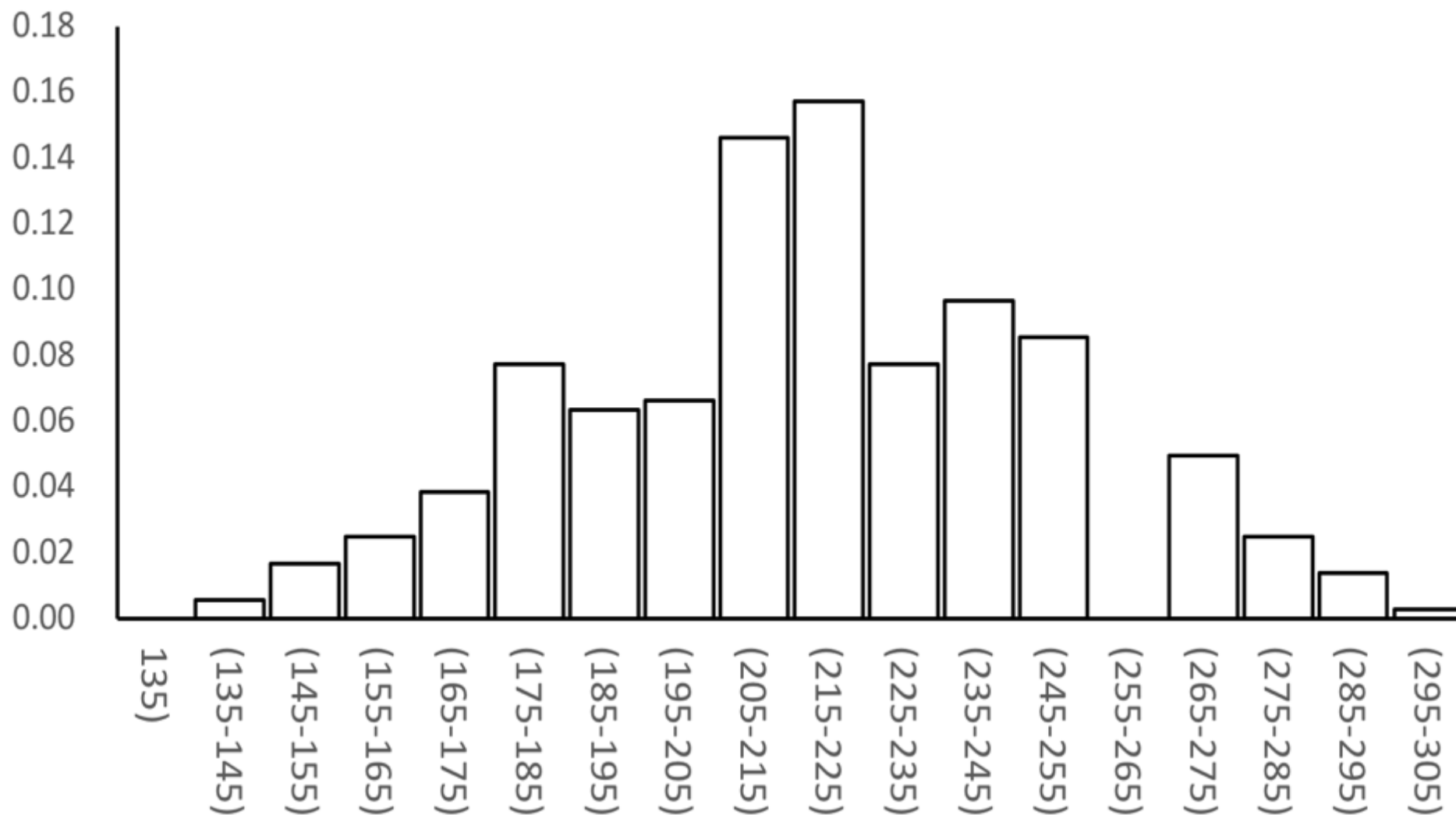
データが欠損してる場合？



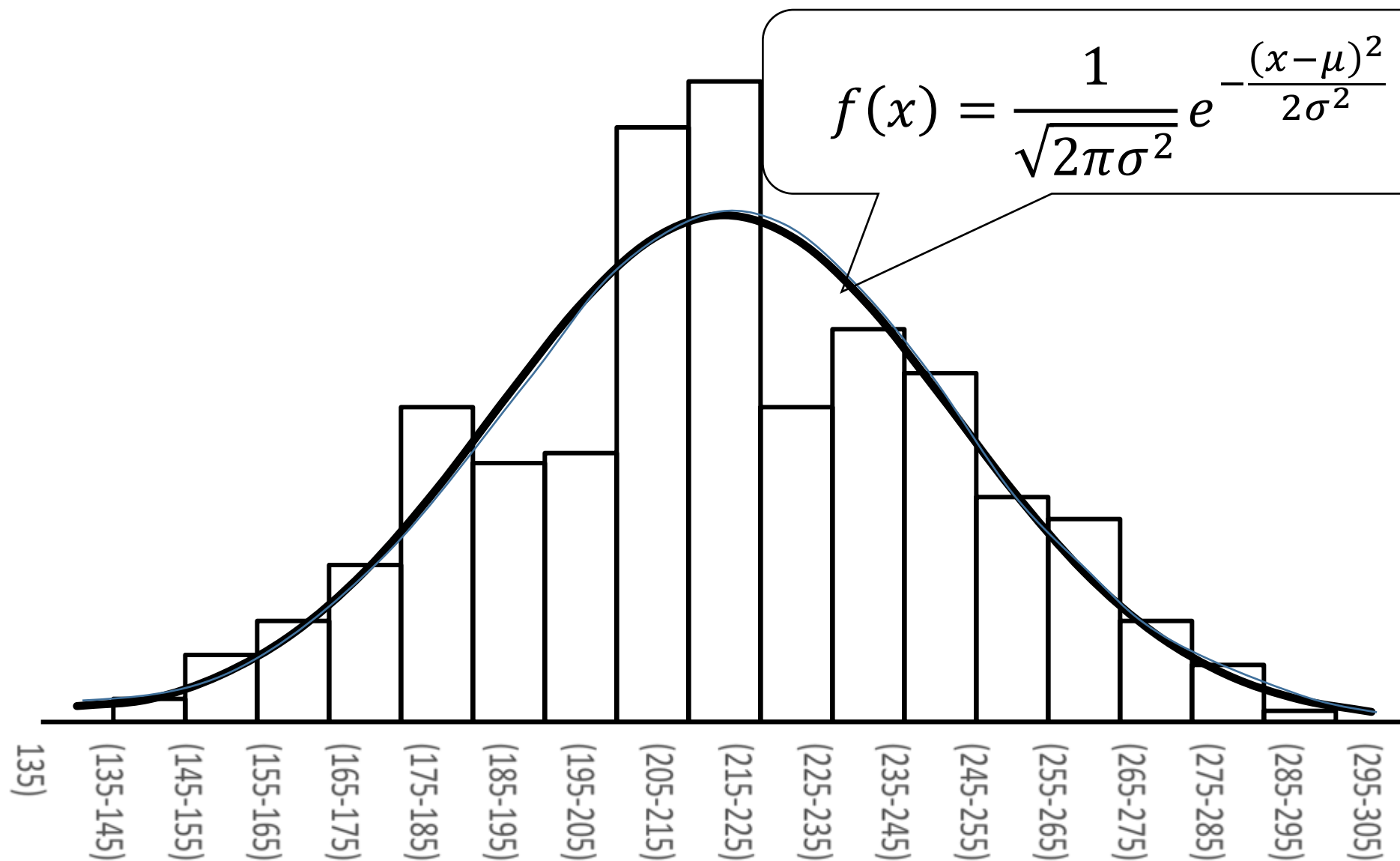
データが欠損してる場合？



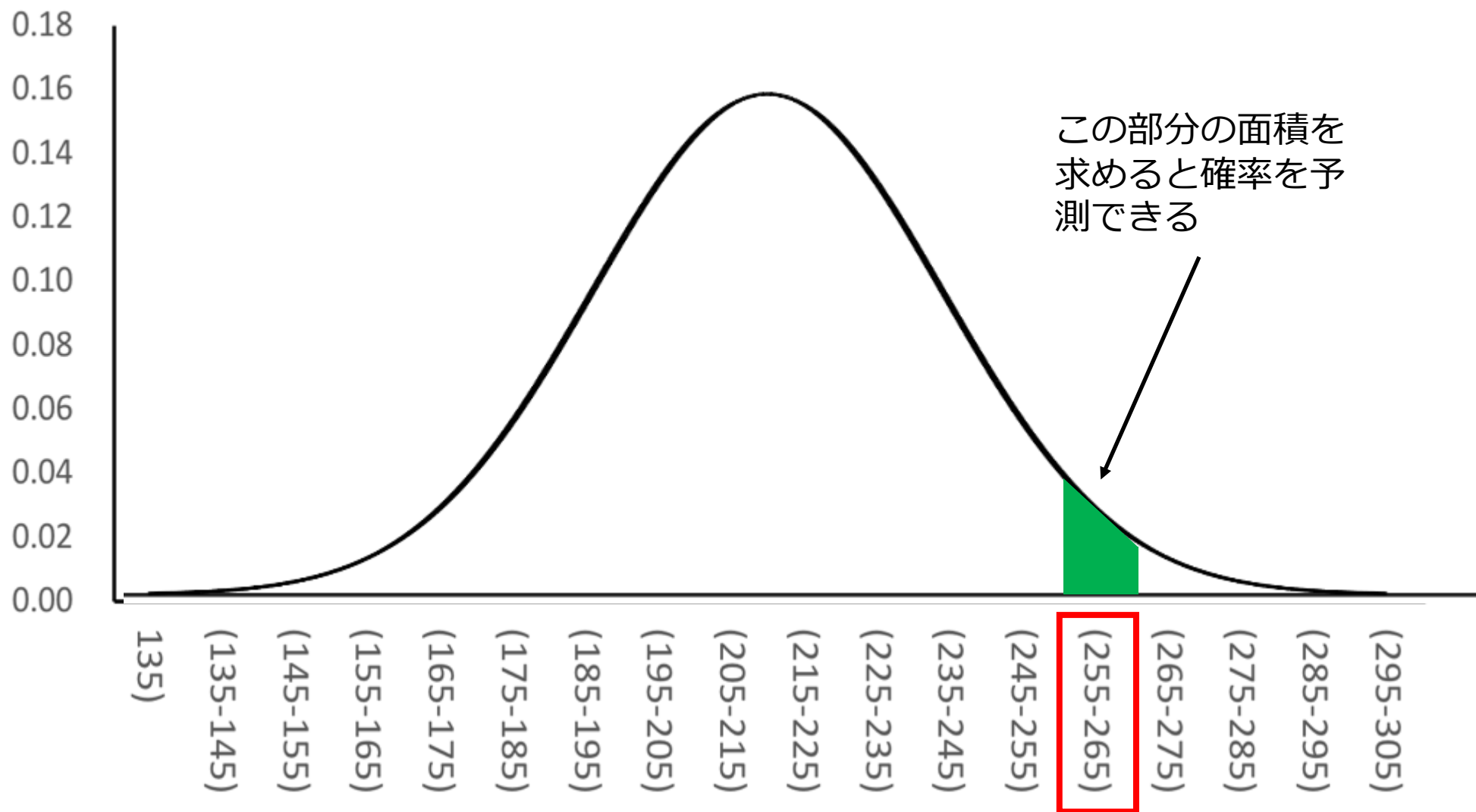
データが欠損してる場合？



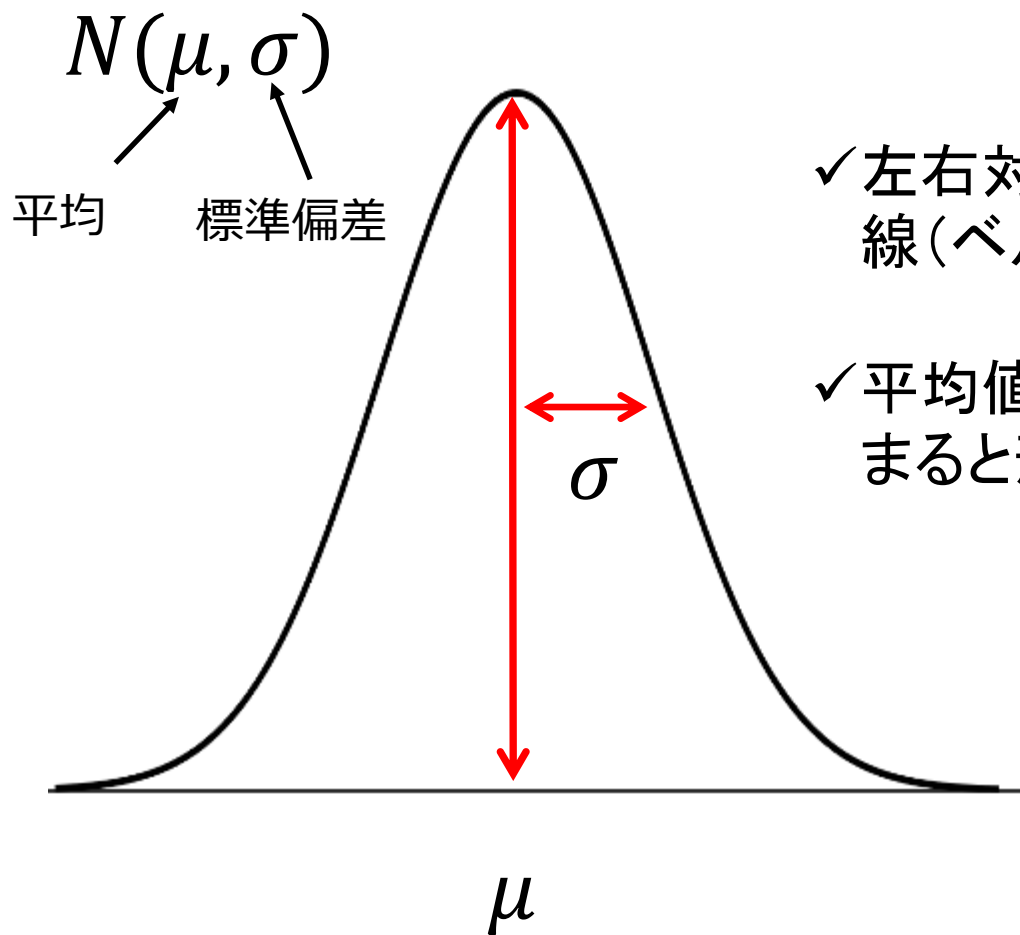
分布を使って近似する



データが欠損してる場合？



正規分布 $N(\mu, \sigma)$

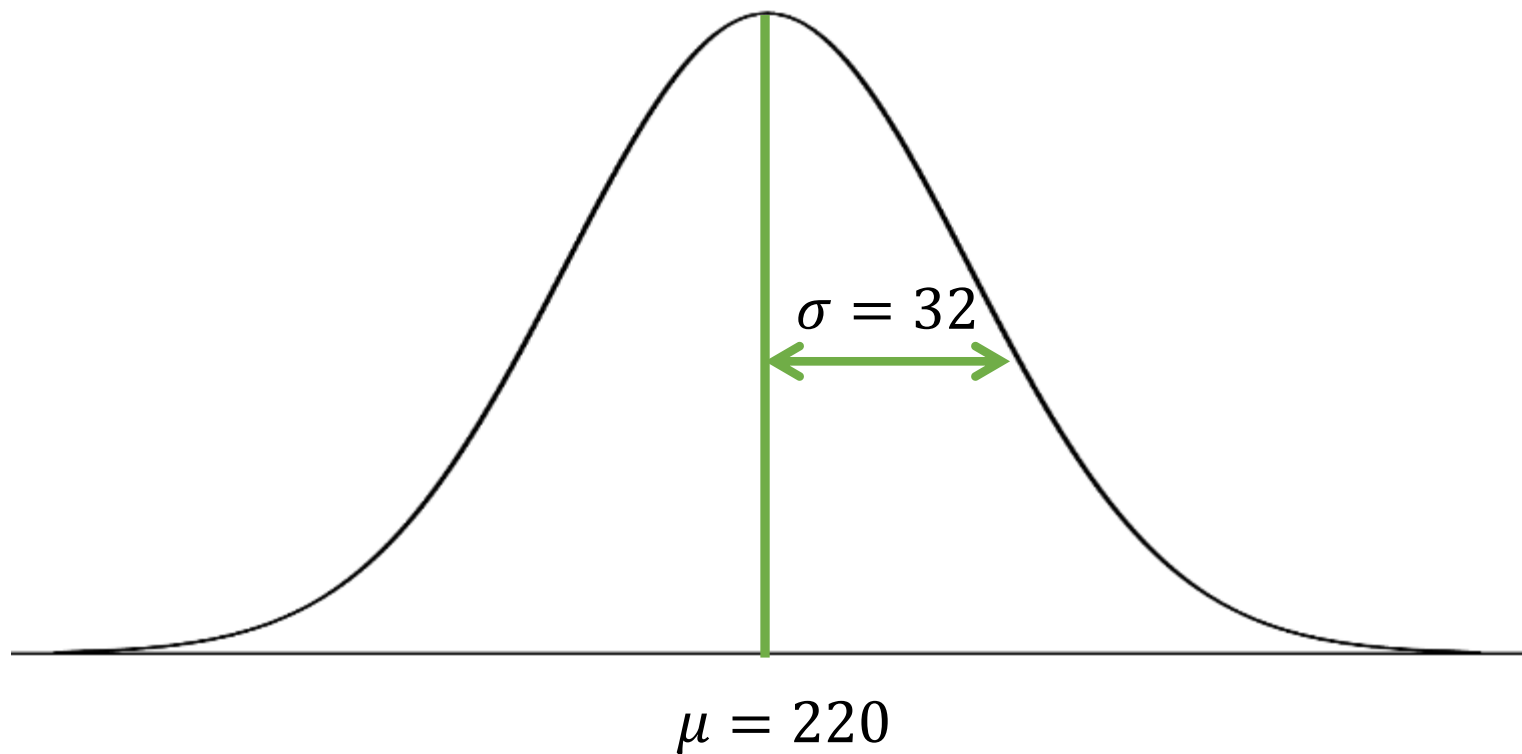


✓ 左右対称になった西洋の釣鐘と似た形状の曲線（ベルカーブ）

✓ 平均値 μ 、標準偏差 σ の2つのパラメータが決まると形が決まる。

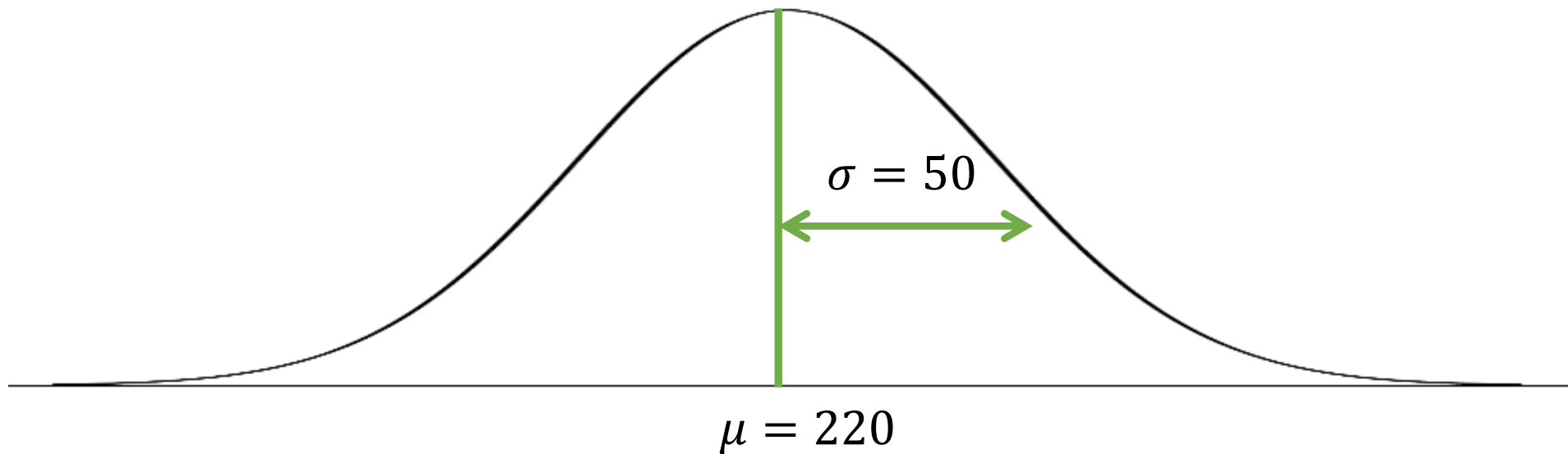
2つのパラメータで形が決まる

$N(220, 32)$



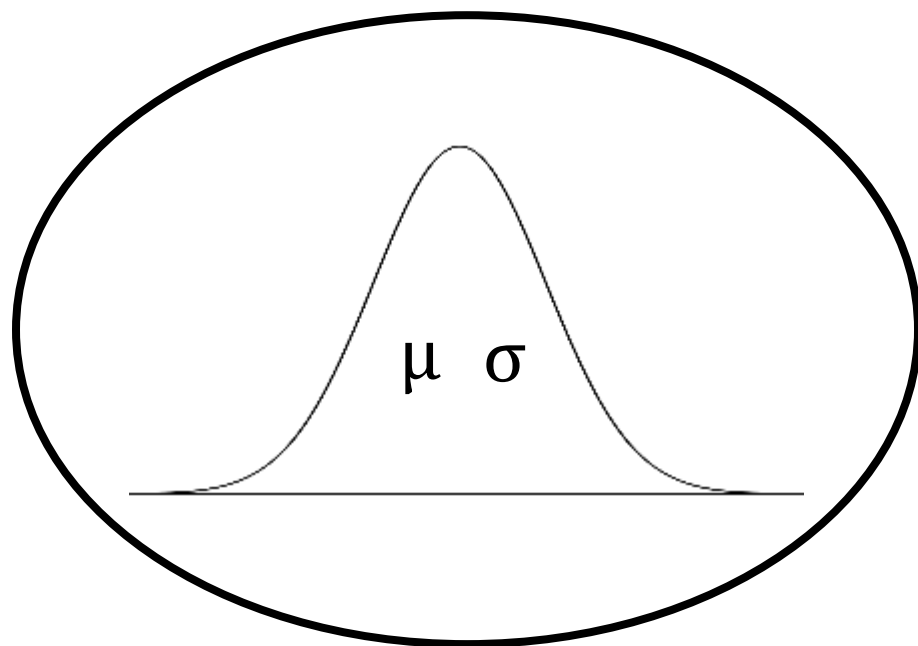
2つのパラメータで形が決まる

$N(220, 50)$

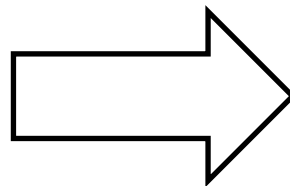


記号の使い方に関するルール

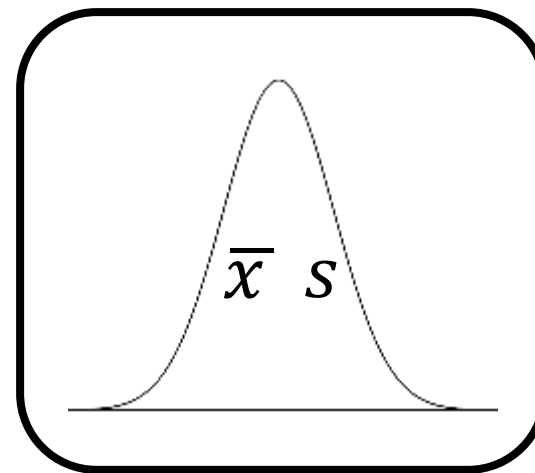
母集団



$N(\mu, \sigma)$



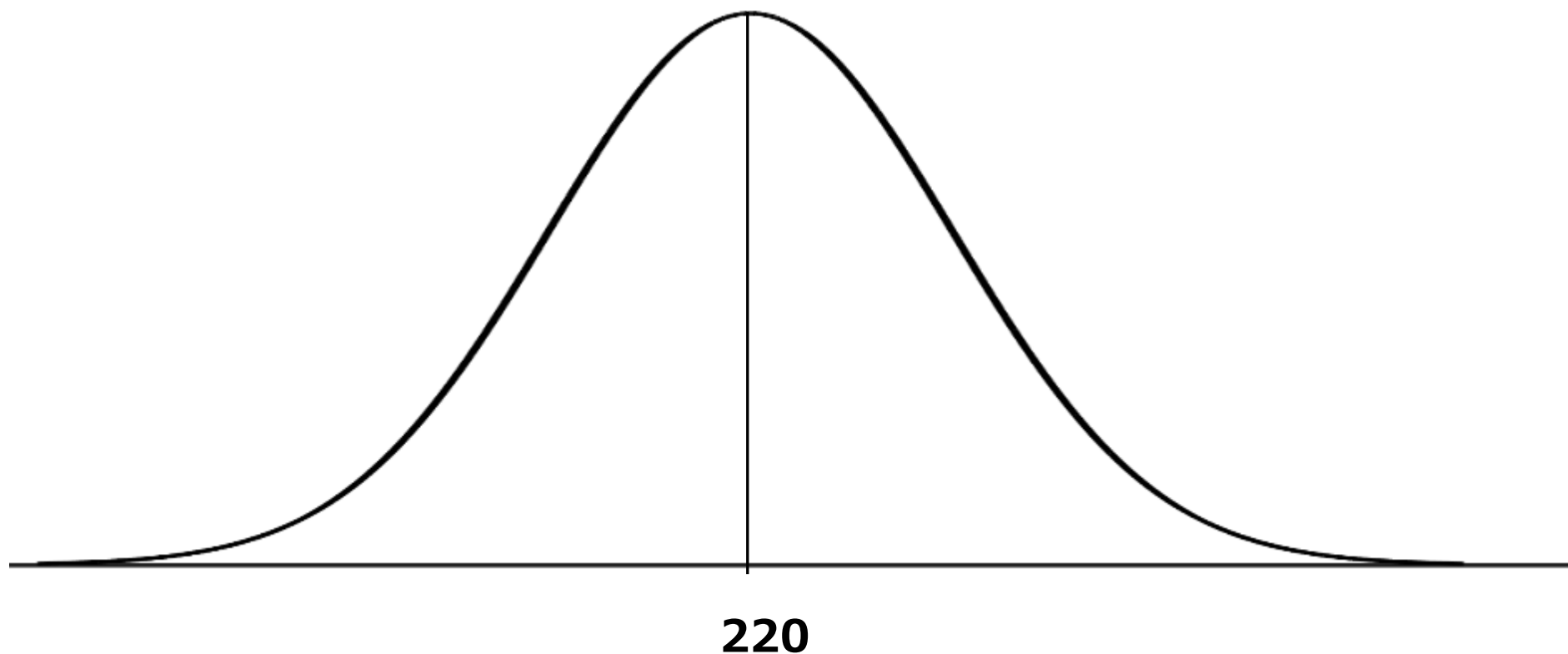
標本



$N(\bar{x}, s)$

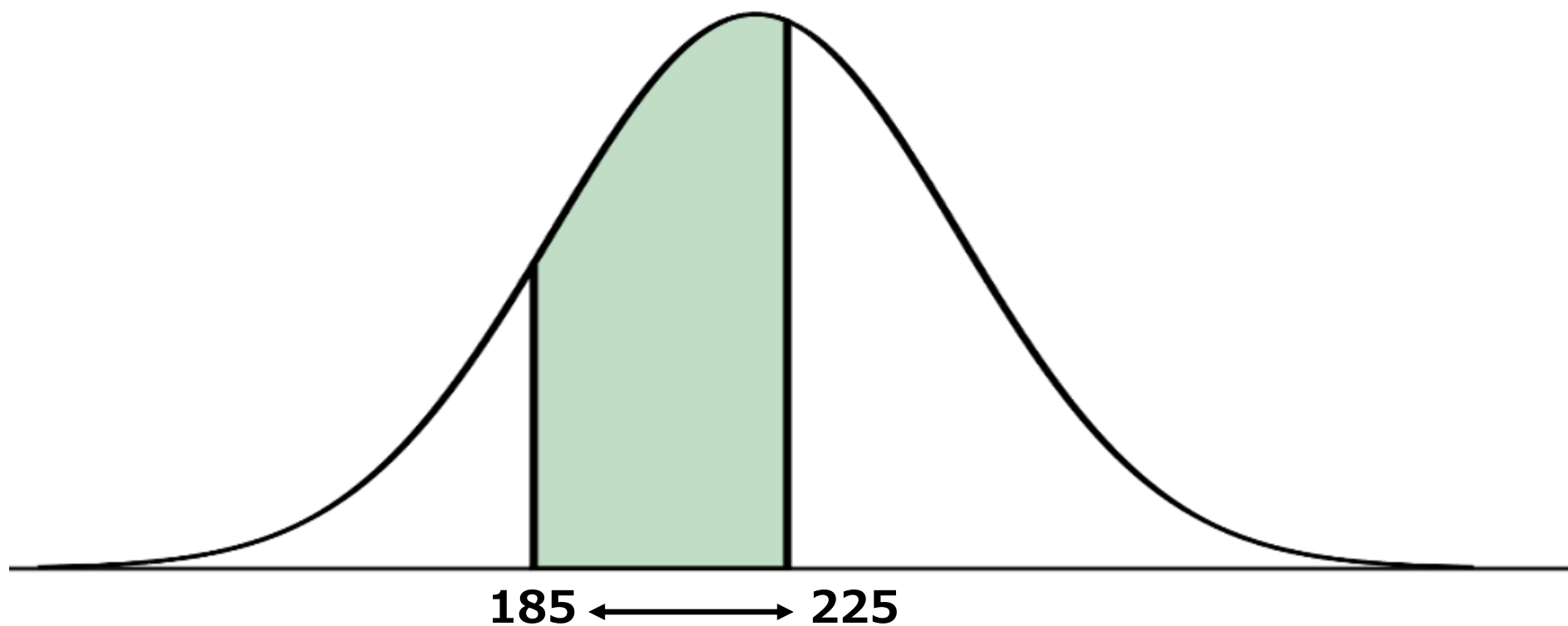
株価が 1 8 5 円から 2 2 5 円の間にある確率は？

$N(220, 32)$ の株価



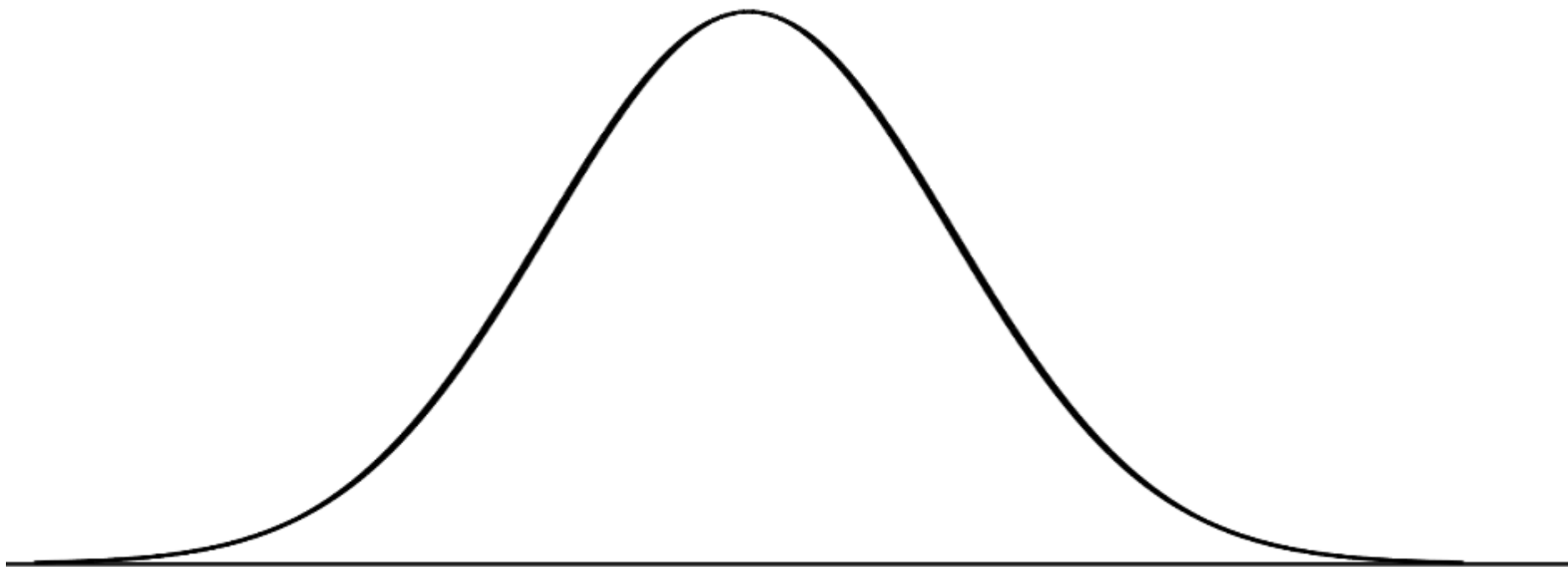
株価が 1 8 5 円から 2 2 5 円の間にある確率は？

$N(220, 32)$ の株価



正規分布の面積の総和は？

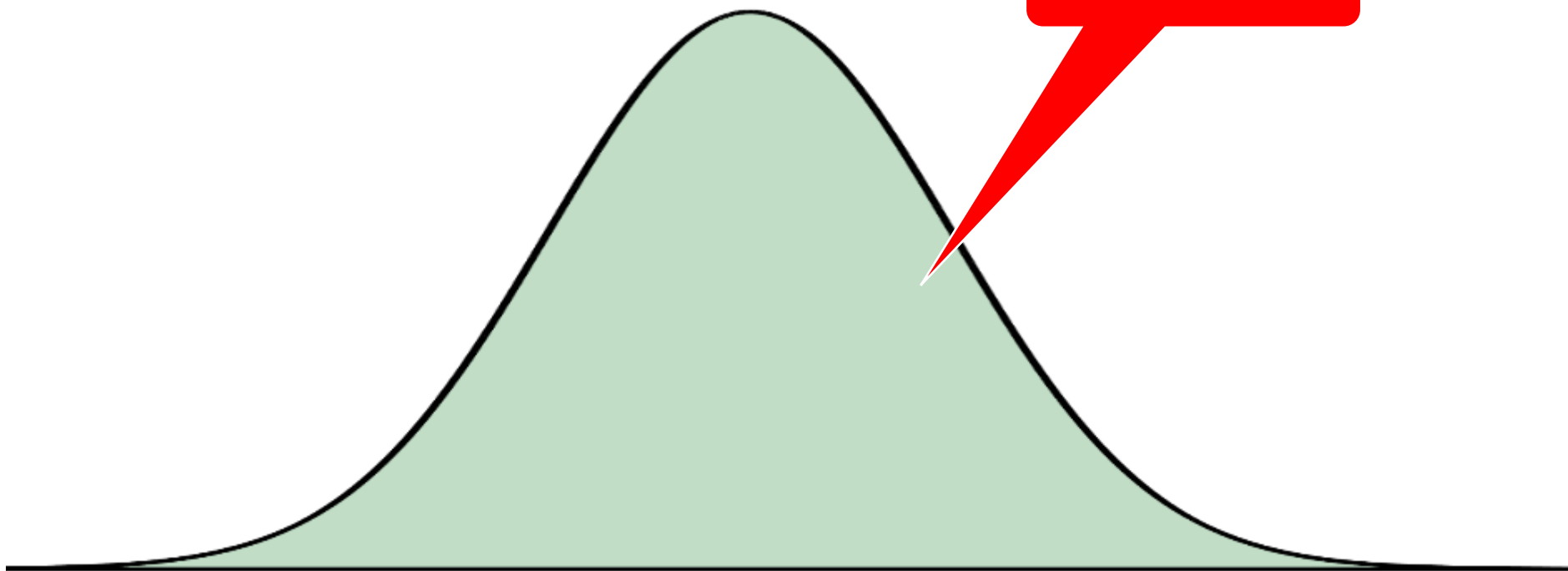
- 面積の総和 = 1



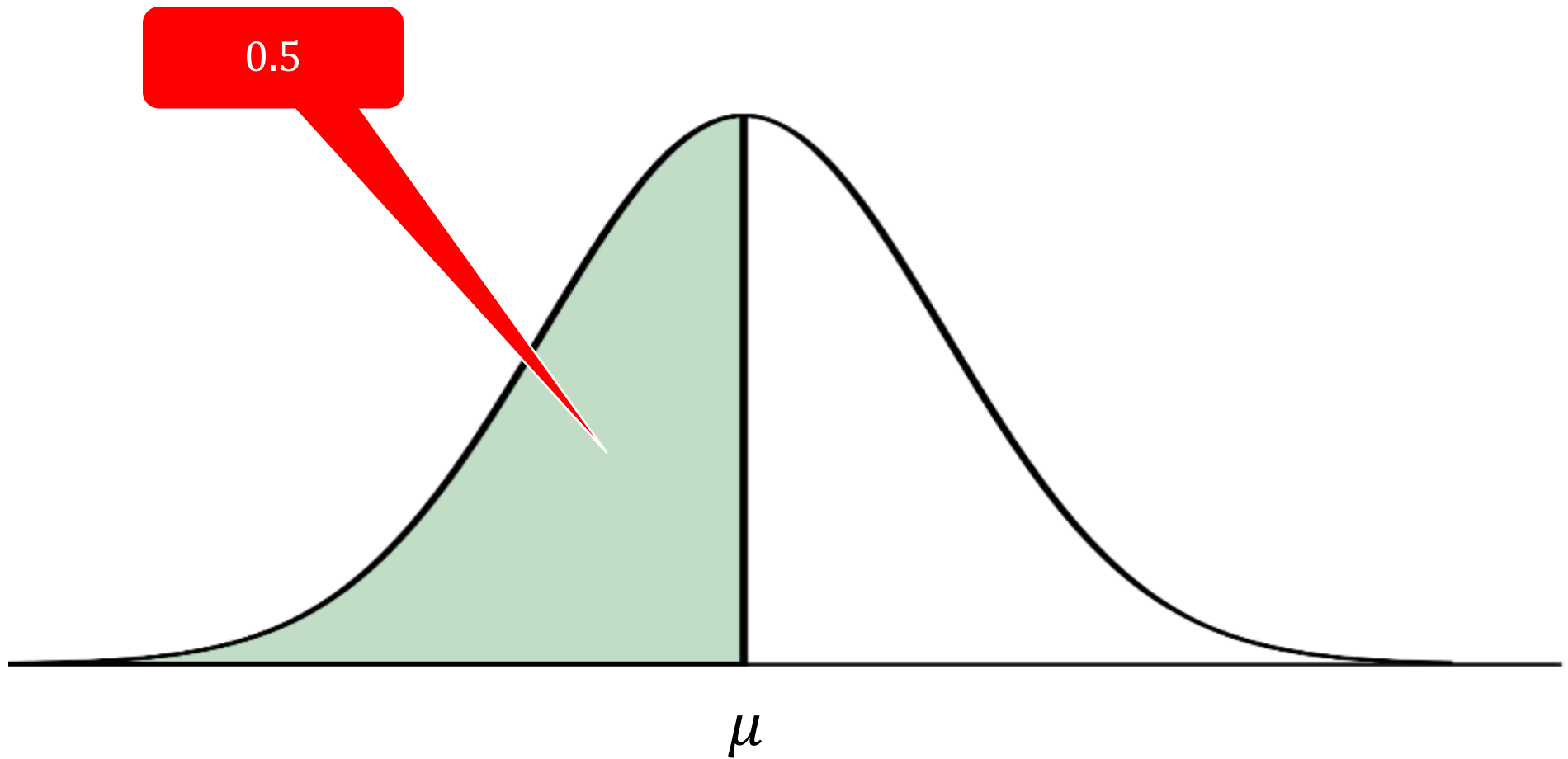
正規分布の面積の総和は？

- 面積の総和 = 1

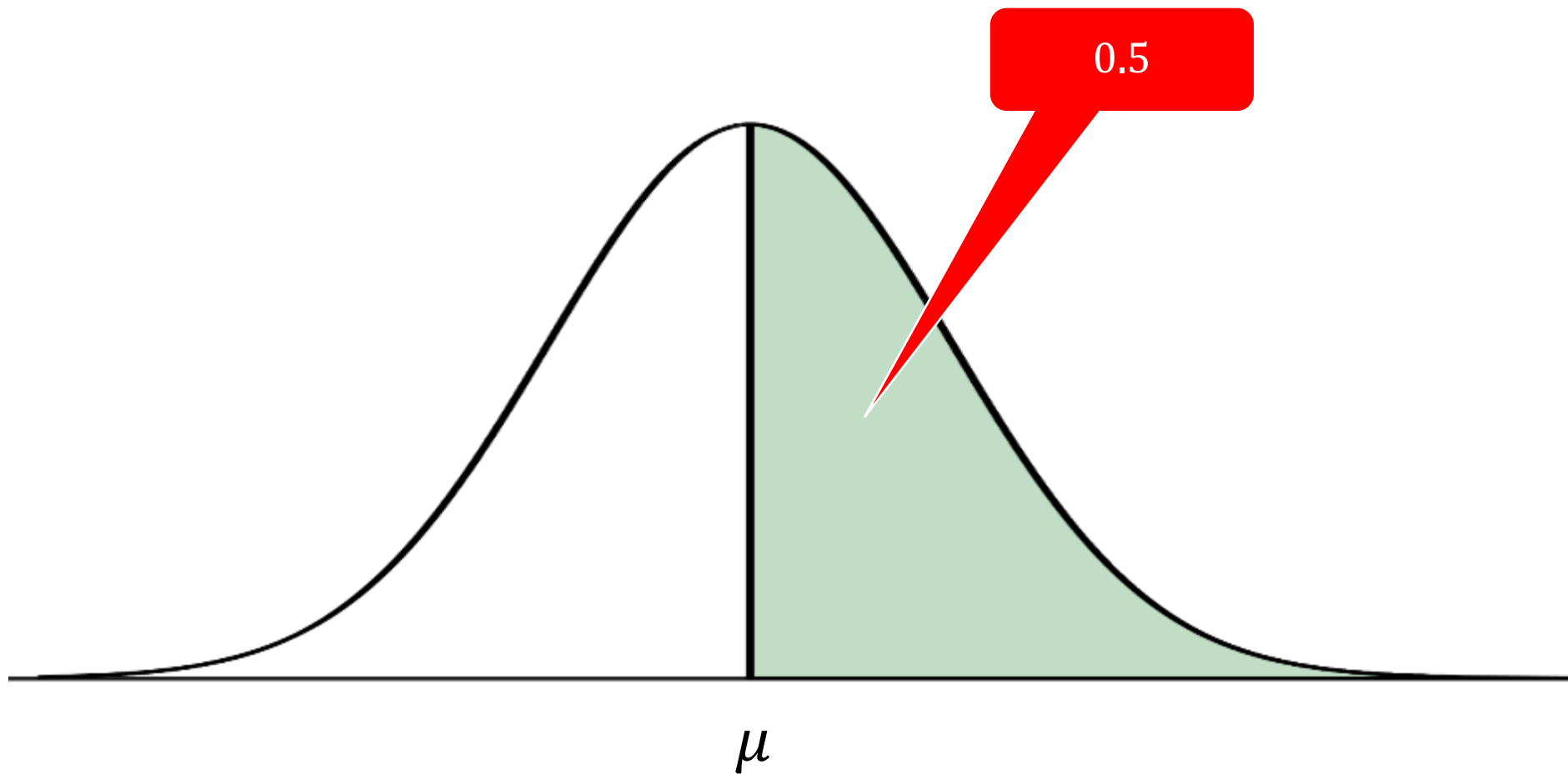
面積の総和 = 1



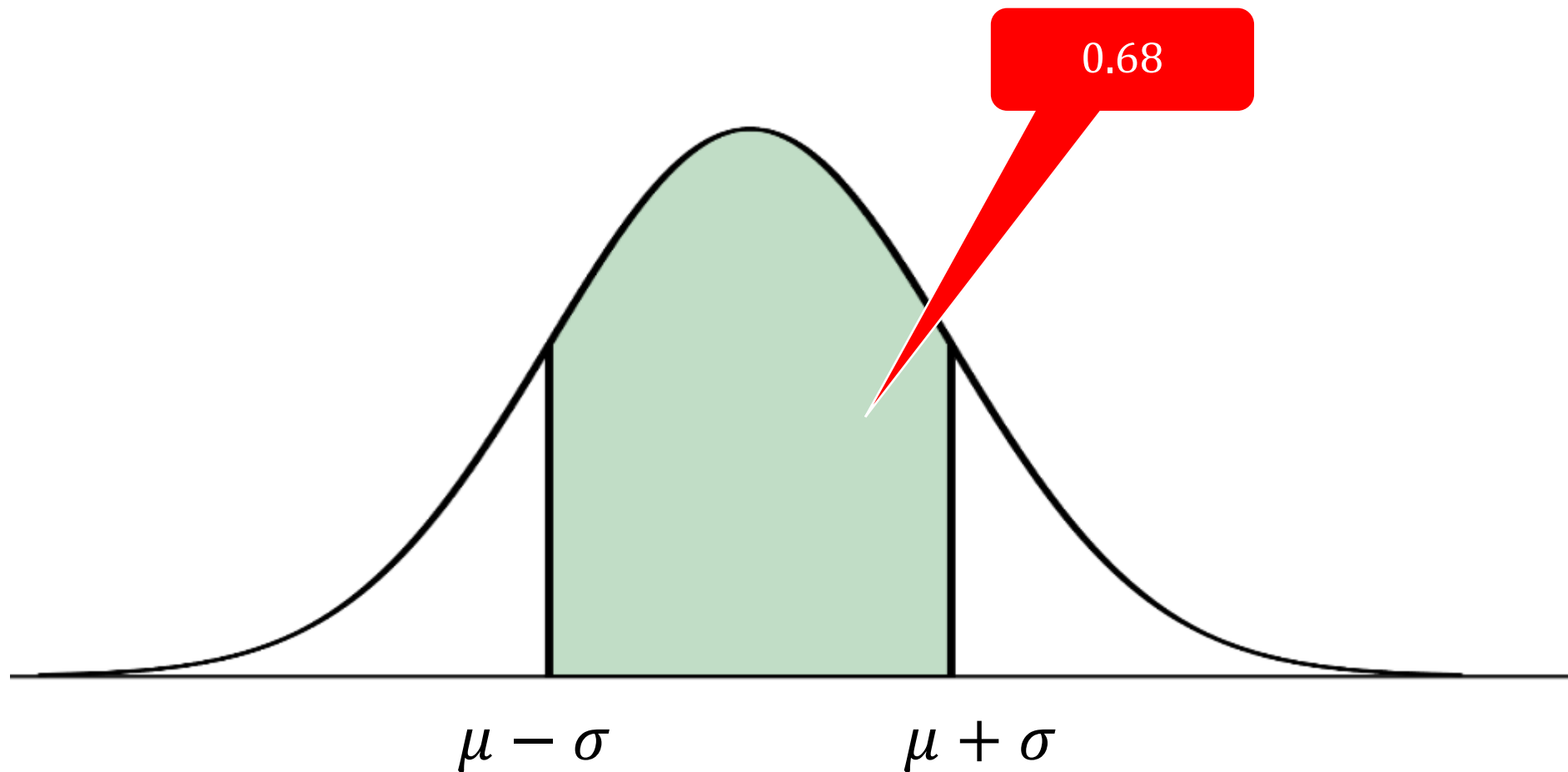
次の面積を求めよ？



次の面積を求めよ？

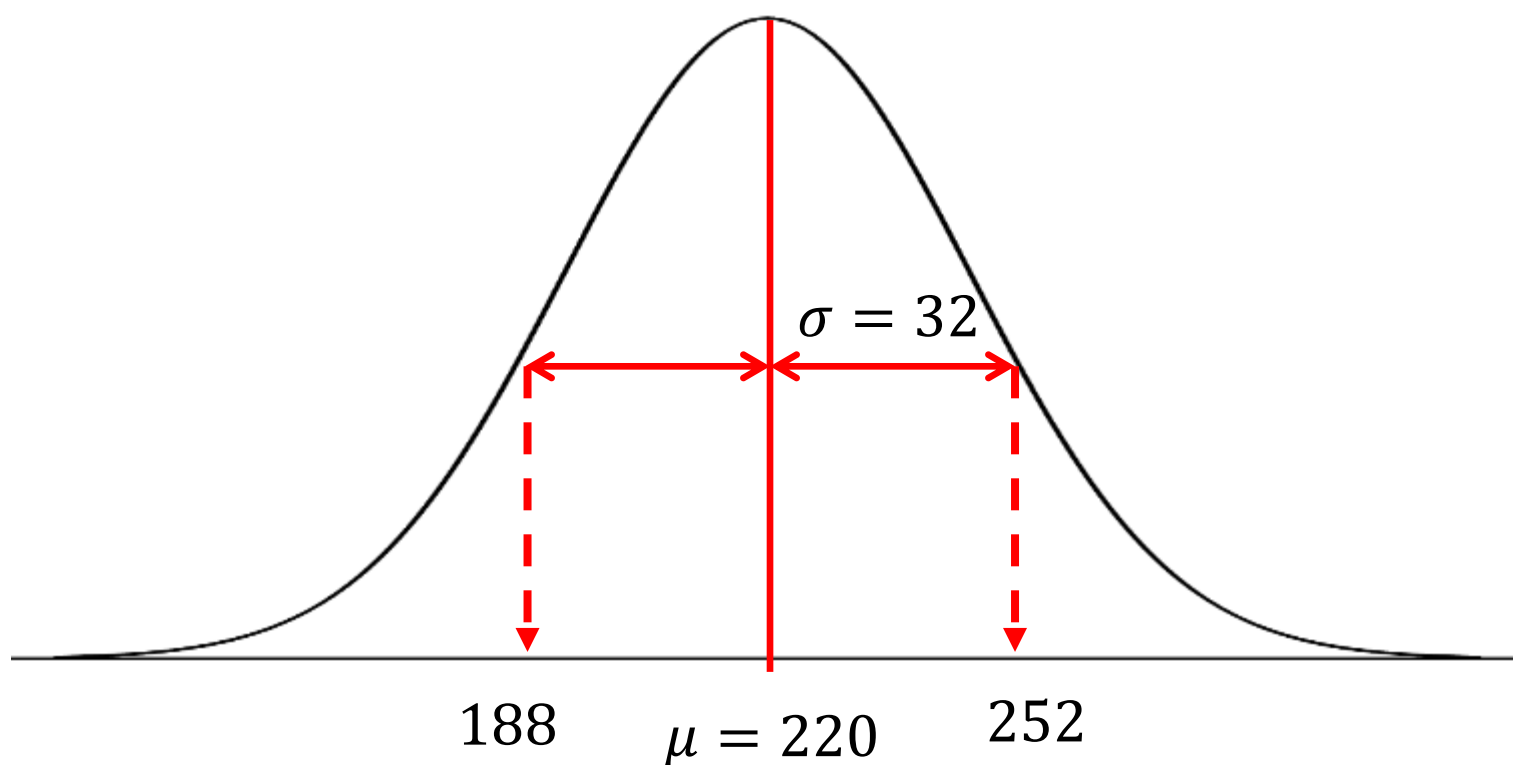


次の面積は？（1σルール）

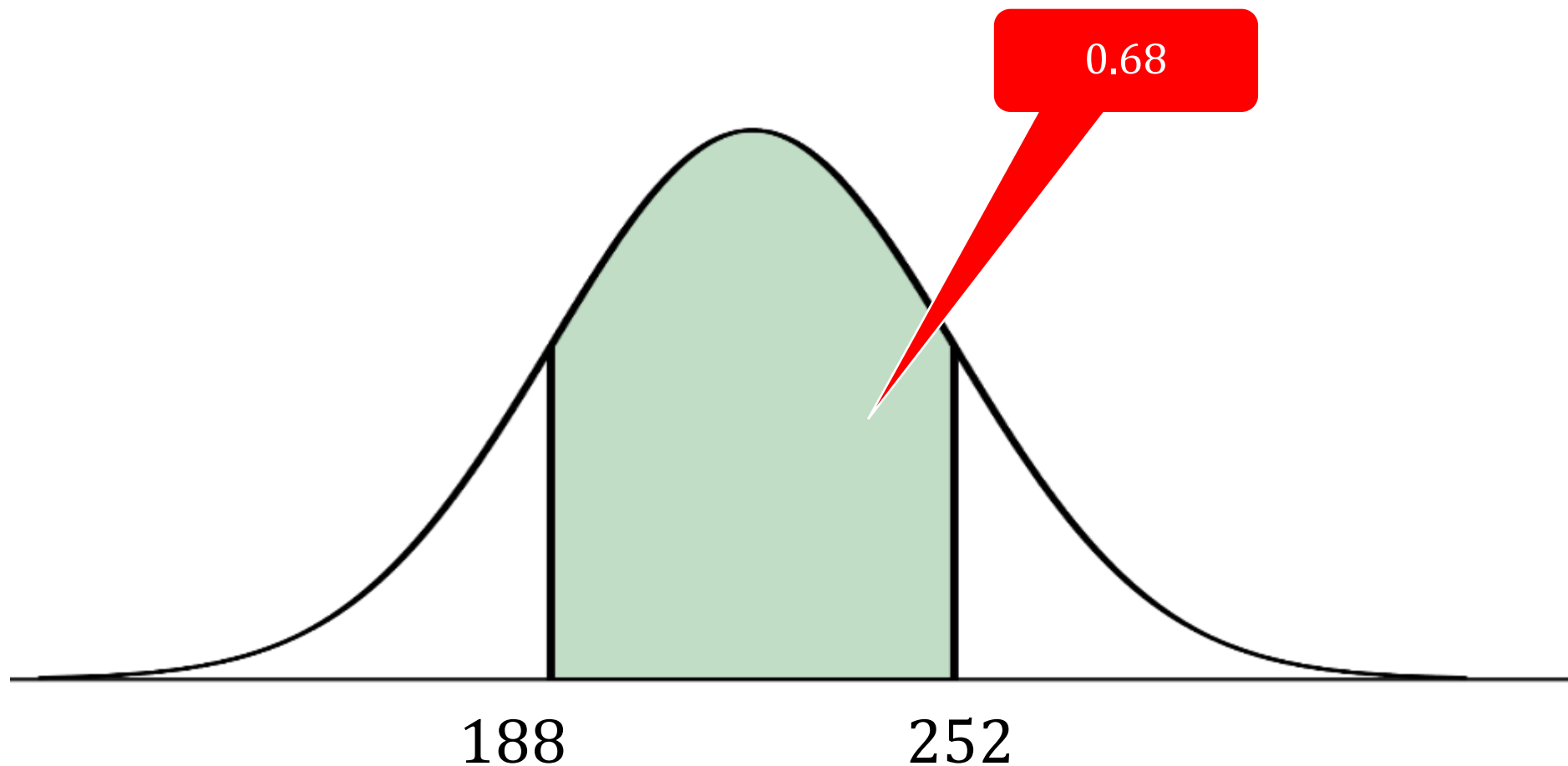


1σルール

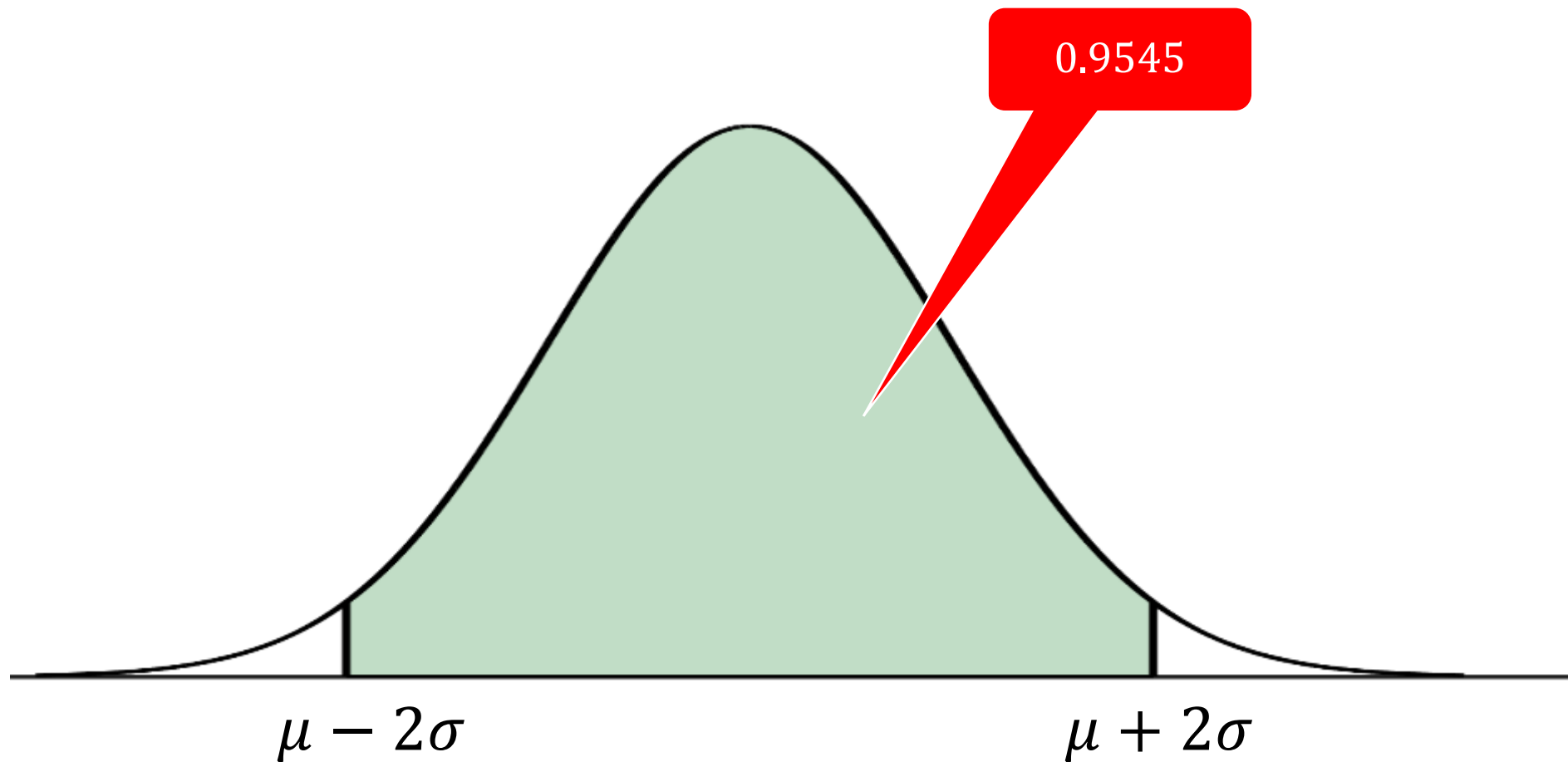
平均 220 円、標準偏差が 32 円で正規分布に株価があるとする。
この株価が 188 円から 252 円を変動する確率は？



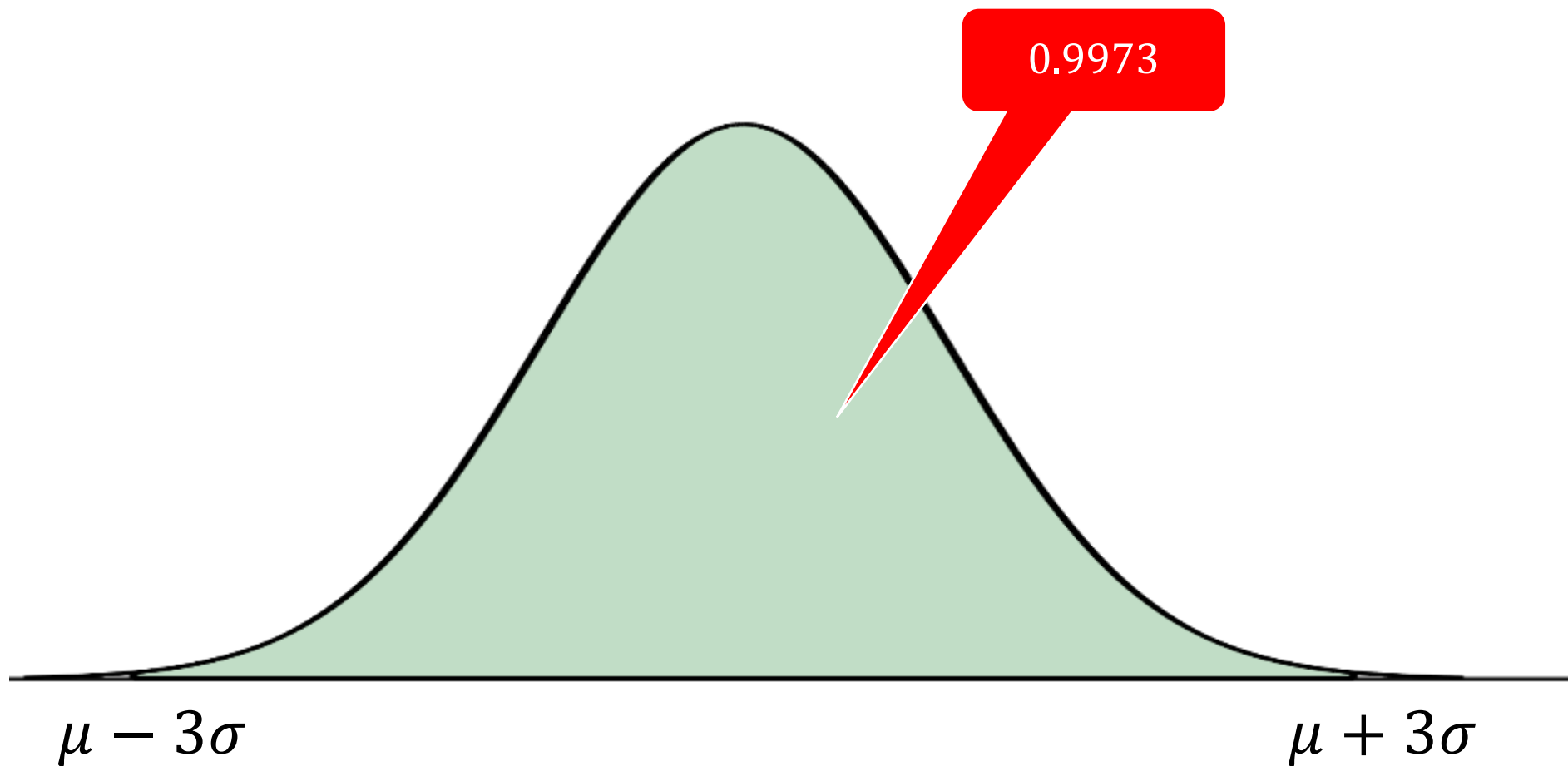
188円から252円で変動する確率は？



次の面積を求めよ？（2σルール）

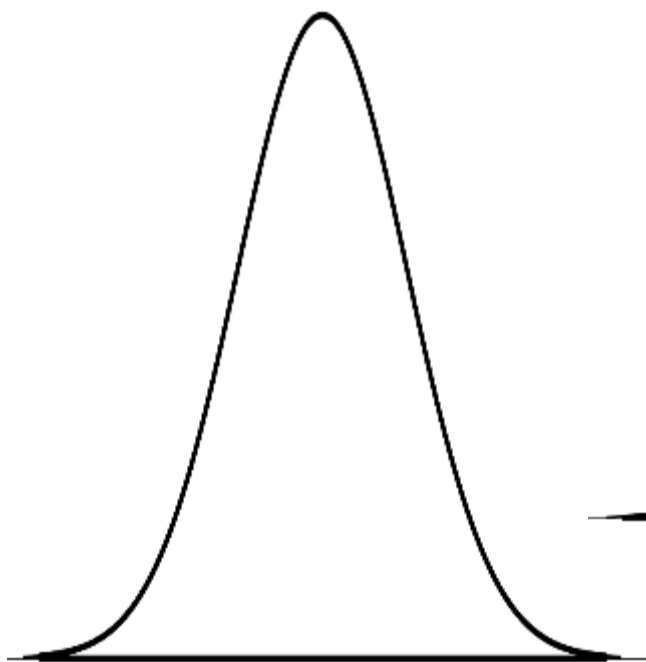


次の面積を求めよ？（3σルール）



無限通りの正規分布 $N(\mu, \sigma)$

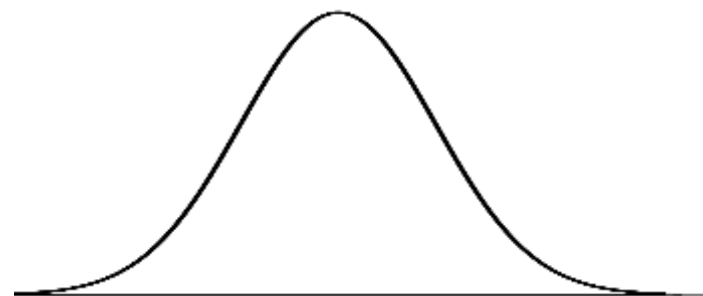
$N(120, 30)$



$N(10, 1)$

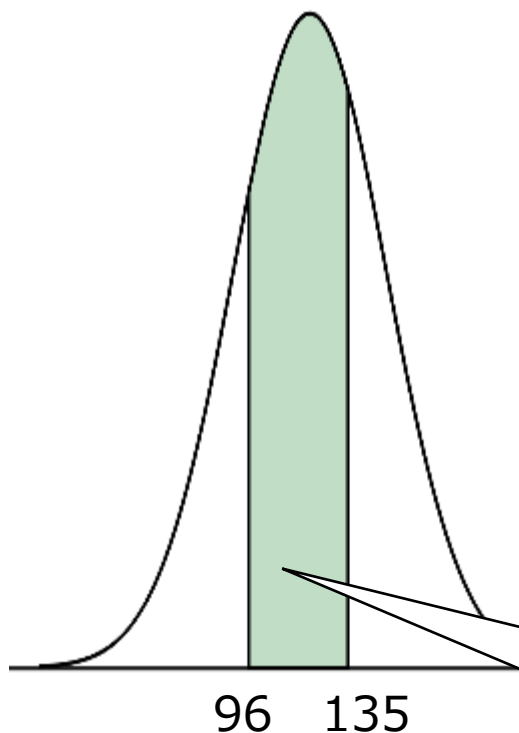


$N(500, 10)$

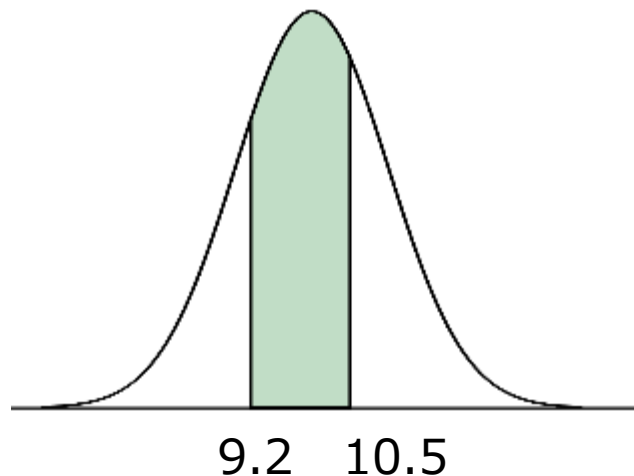


無限通りの正規分布 $N(\mu, \sigma)$

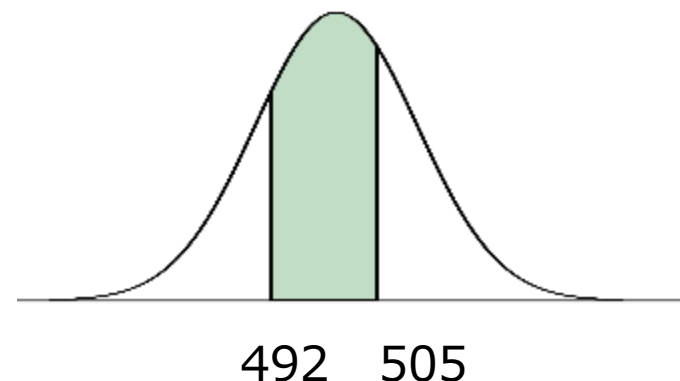
$N(120, 30)$



$N(10, 1)$



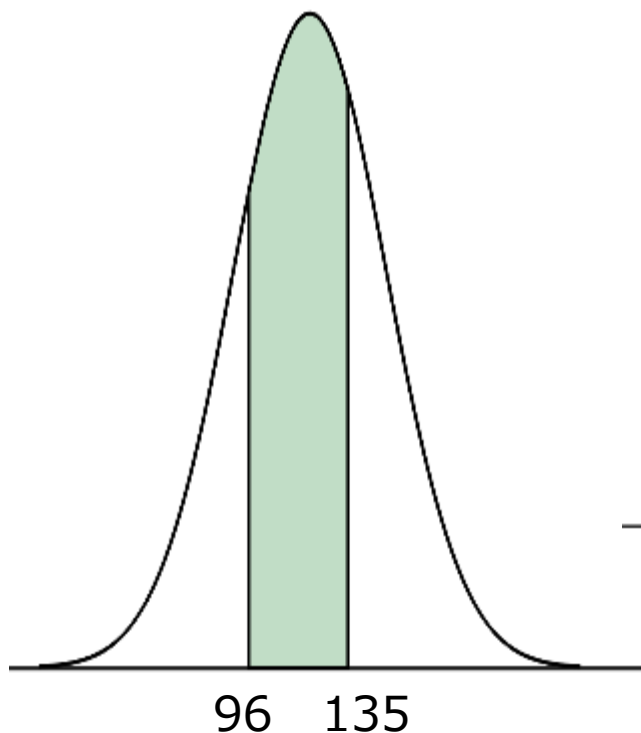
$N(500, 10)$



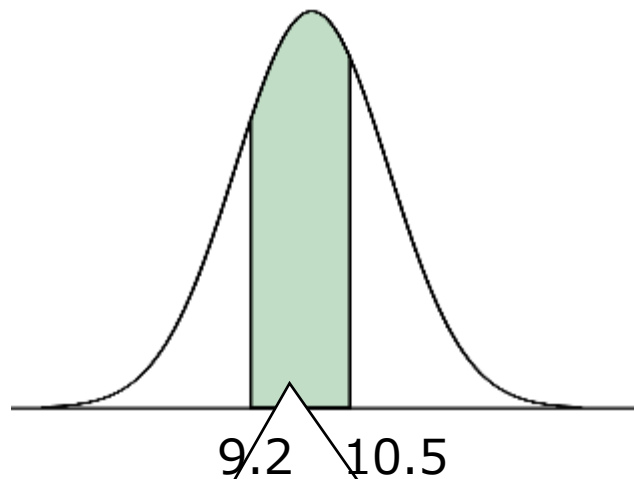
$$P(96 < X < 135) = \int_{96}^{135} \frac{1}{\sqrt{2\pi}30^2} \exp\left(-\frac{(x-120)^2}{2 \times 30^2}\right) dx$$

無限通りの正規分布 $N(\mu, \sigma)$

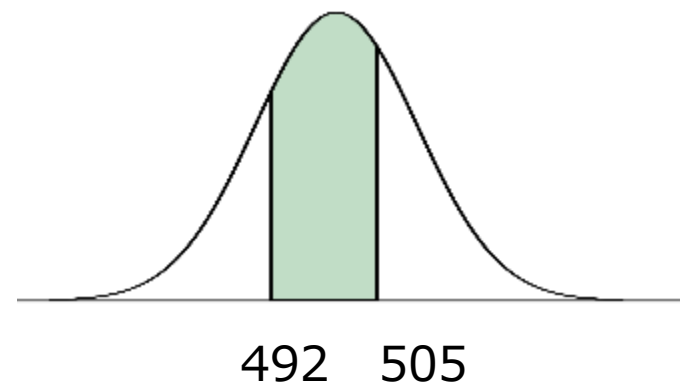
$N(120, 30)$



$N(10, 1)$



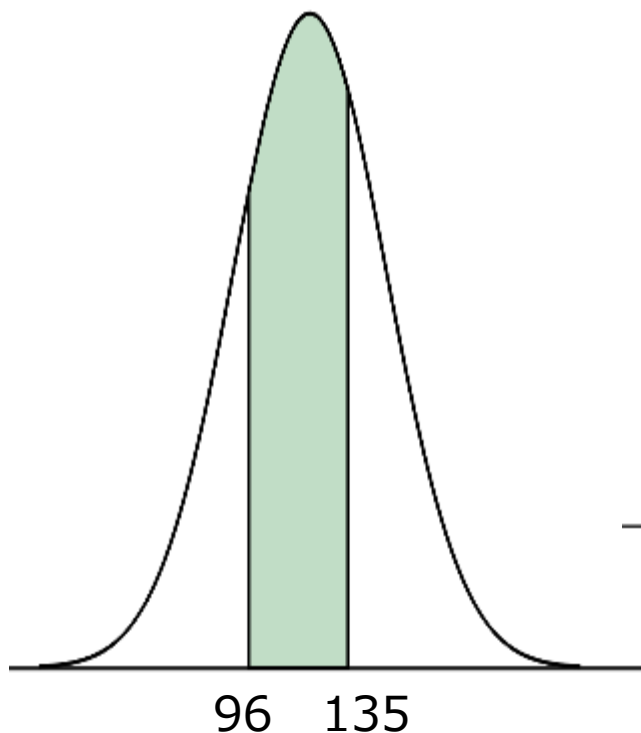
$N(500, 10)$



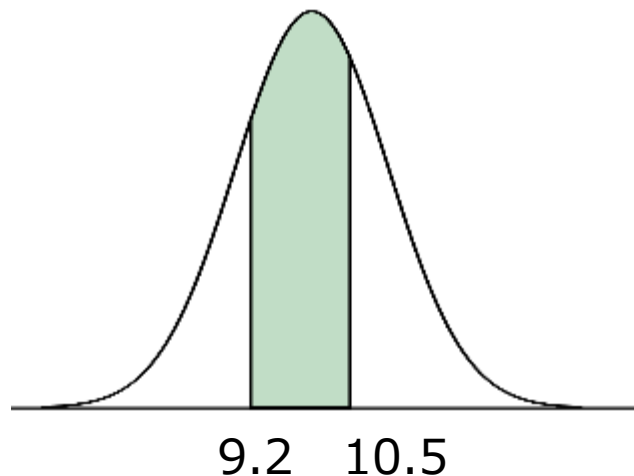
$$P(9.2 < X < 10.5) = \int_{9.2}^{10.5} \frac{1}{\sqrt{2\pi}10^2} \exp\left(-\frac{(x-10)^2}{2 \times 10.5^2}\right) dx$$

無限通りの正規分布 $N(\mu, \sigma)$

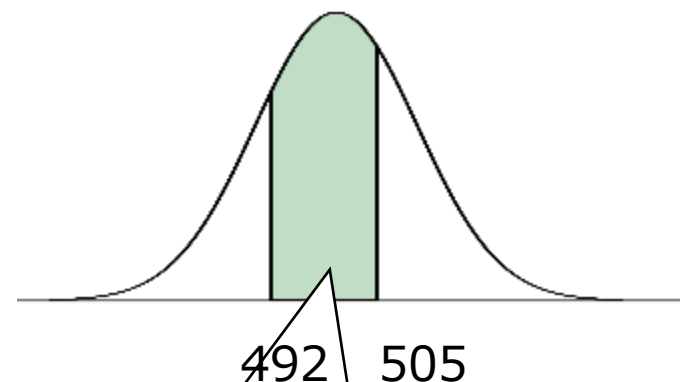
$N(120, 30)$



$N(10, 1)$



$N(500, 10)$



$$P(492 < X < 505) = \int_{492}^{505} \frac{1}{\sqrt{2\pi 500^2}} \exp\left(-\frac{(x - 500)^2}{2 \times 10^2}\right) dx$$

一つの正規分布に統一する

$N(120,30)$

$N(10,1)$

$N(500,10)$

$$Z = \frac{x - \mu}{\sigma}$$

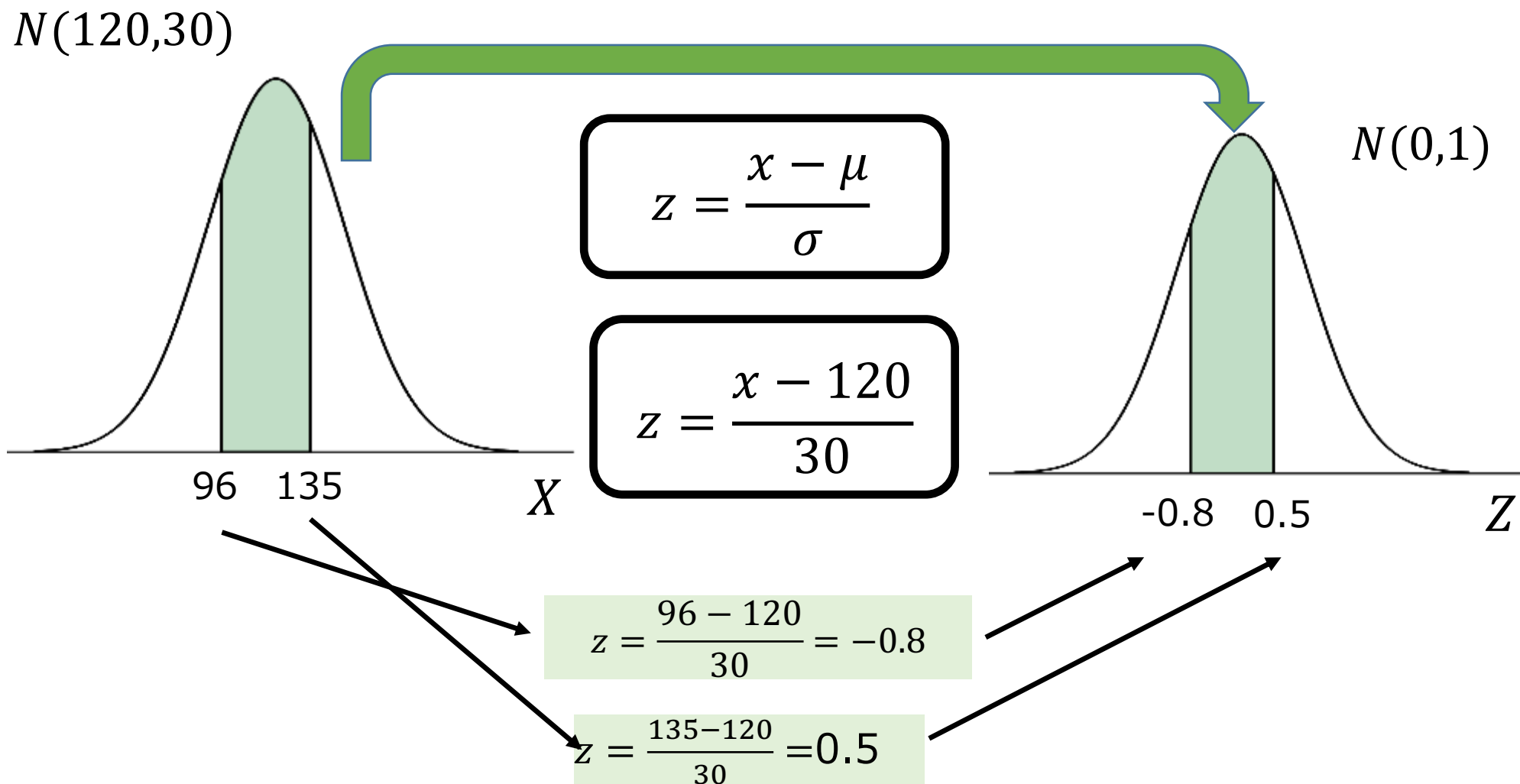
変換公式

標準化正規分布

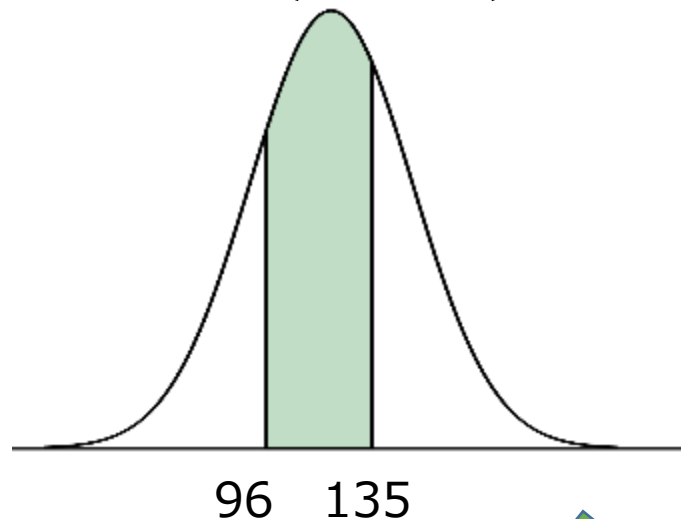
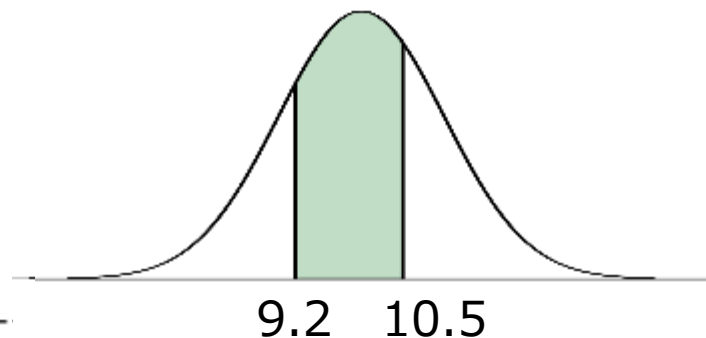
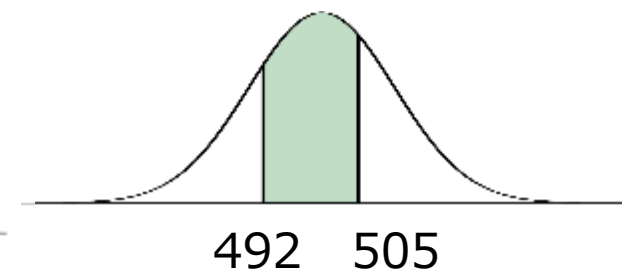
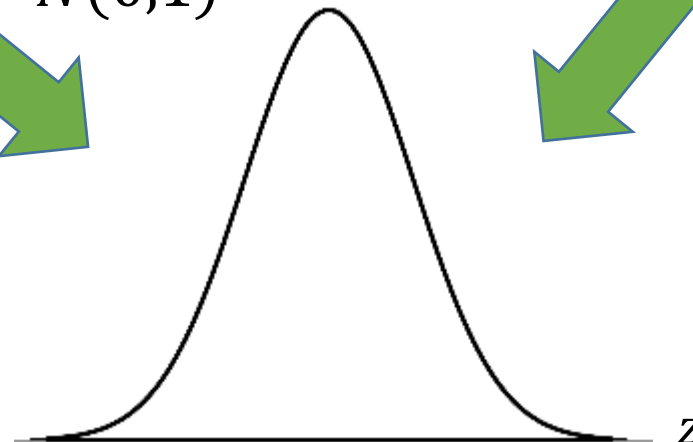
$N(0,1)$

Z

標準化正規分布に変換する



一つの正規分布に統一する

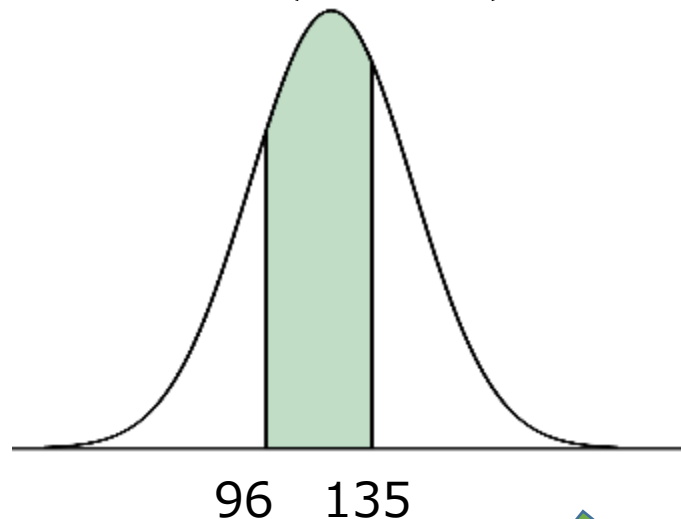
 $N(120, 30)$  $N(10, 1)$  $N(500, 10)$  $N(0, 1)$ 

$$Z = \frac{x - \mu}{\sigma}$$

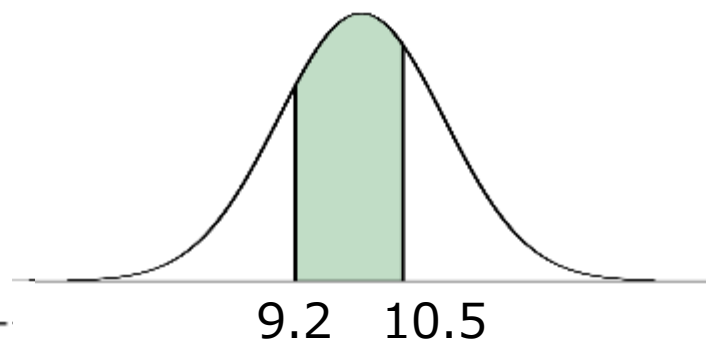
変換公式

一つの正規分布に統一する

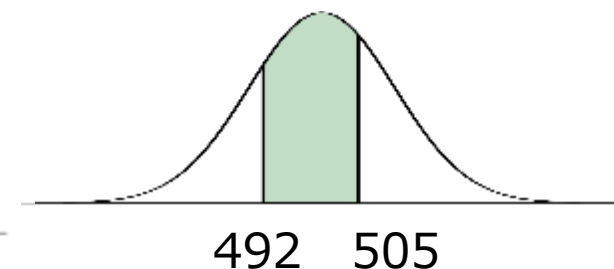
$N(120,30)$



$N(10,1)$



$N(500,10)$

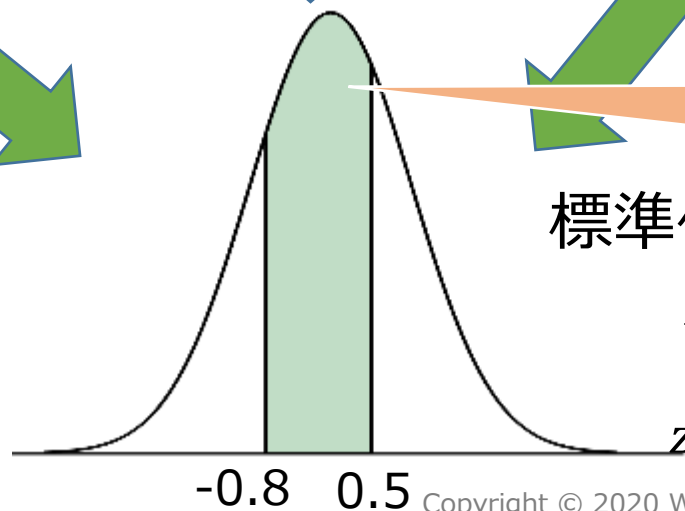


$$Z = \frac{x - \mu}{\sigma}$$

変換公式

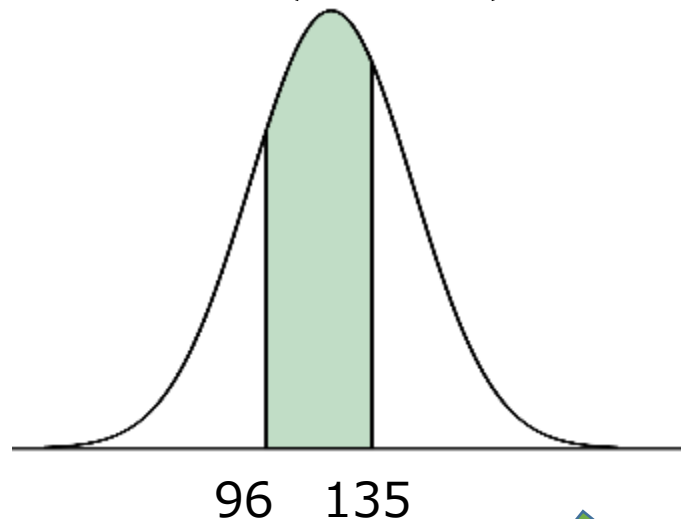
$N(0,1)$
の面積は既知

標準化正規分布
 $N(0,1)$

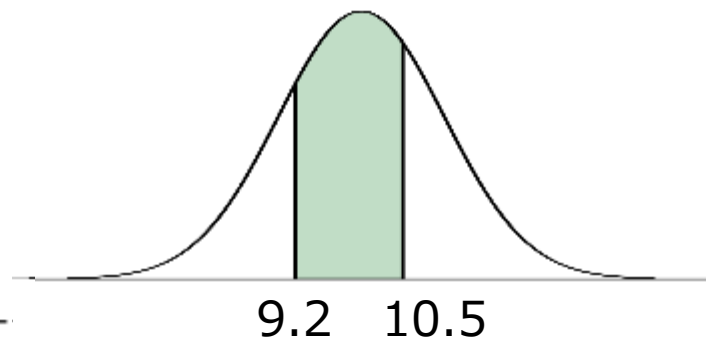


一つの正規分布に統一する

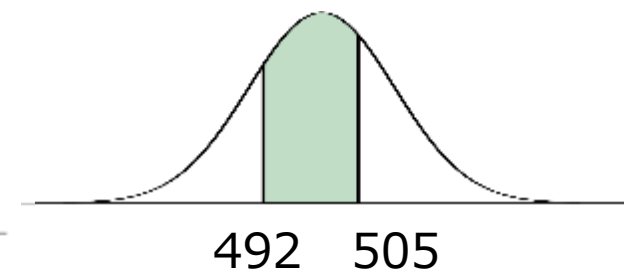
$N(120,30)$



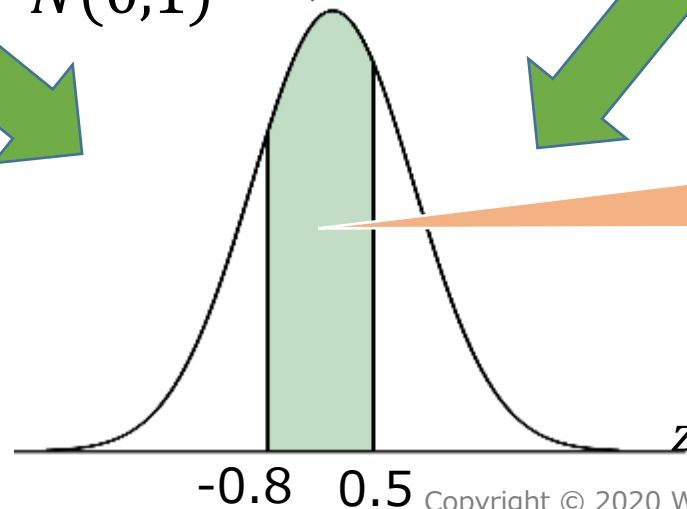
$N(10,1)$



$N(500,10)$



$N(0,1)$



$$z = \frac{x - \mu}{\sigma}$$

変換公式

Z表から
面積 = 0.48

Z表を使って確率を求める

$N(220,32)$ の正規分布において、255以下の確率を求めよ

$N(220,32)$

255

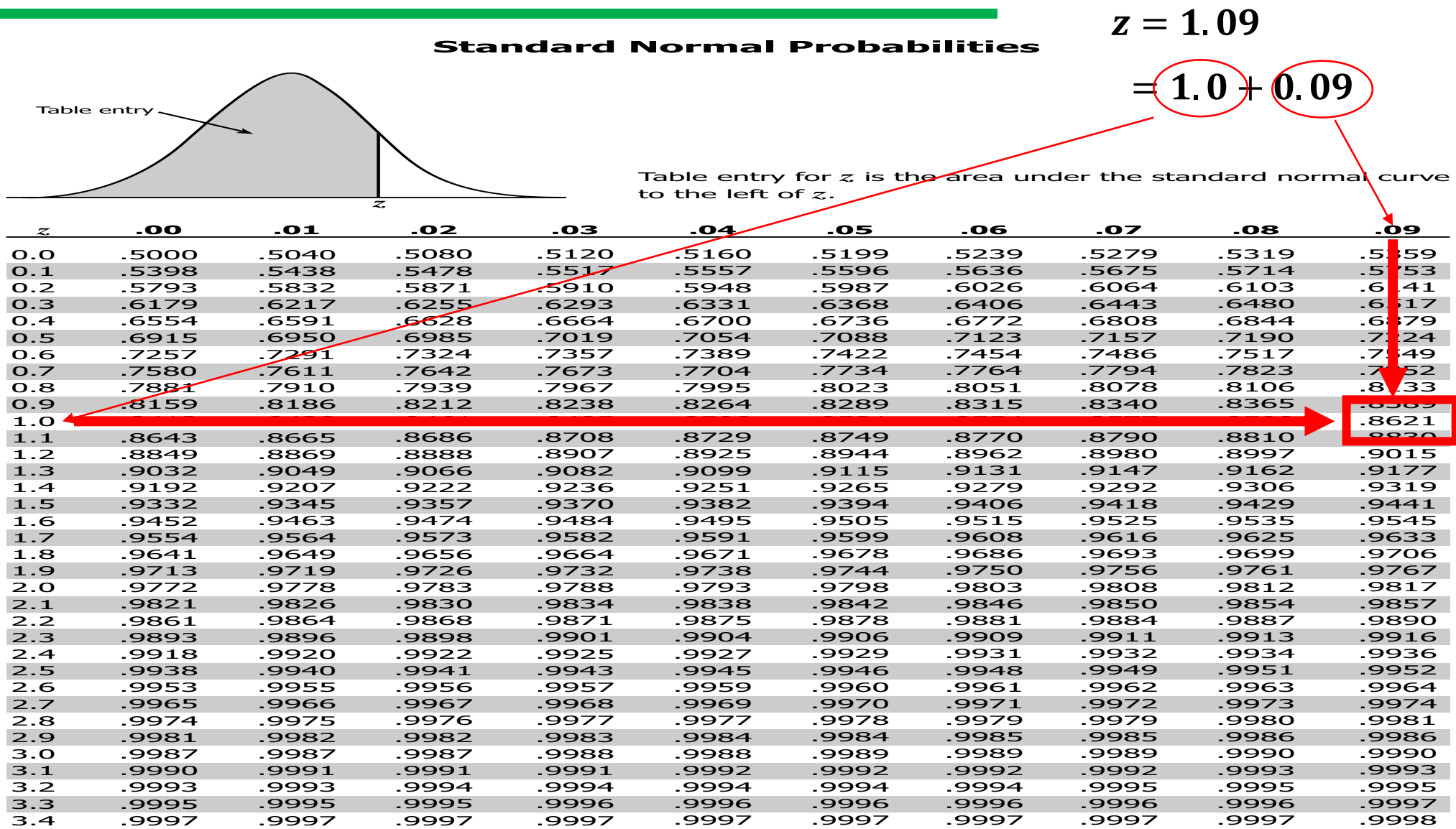
255を $N(0,1)$ の点に変換 (z値を求める)

$$z = \frac{x - \mu}{\sigma} = \frac{255 - 220}{32} = 1.09$$

$N(0,1)$

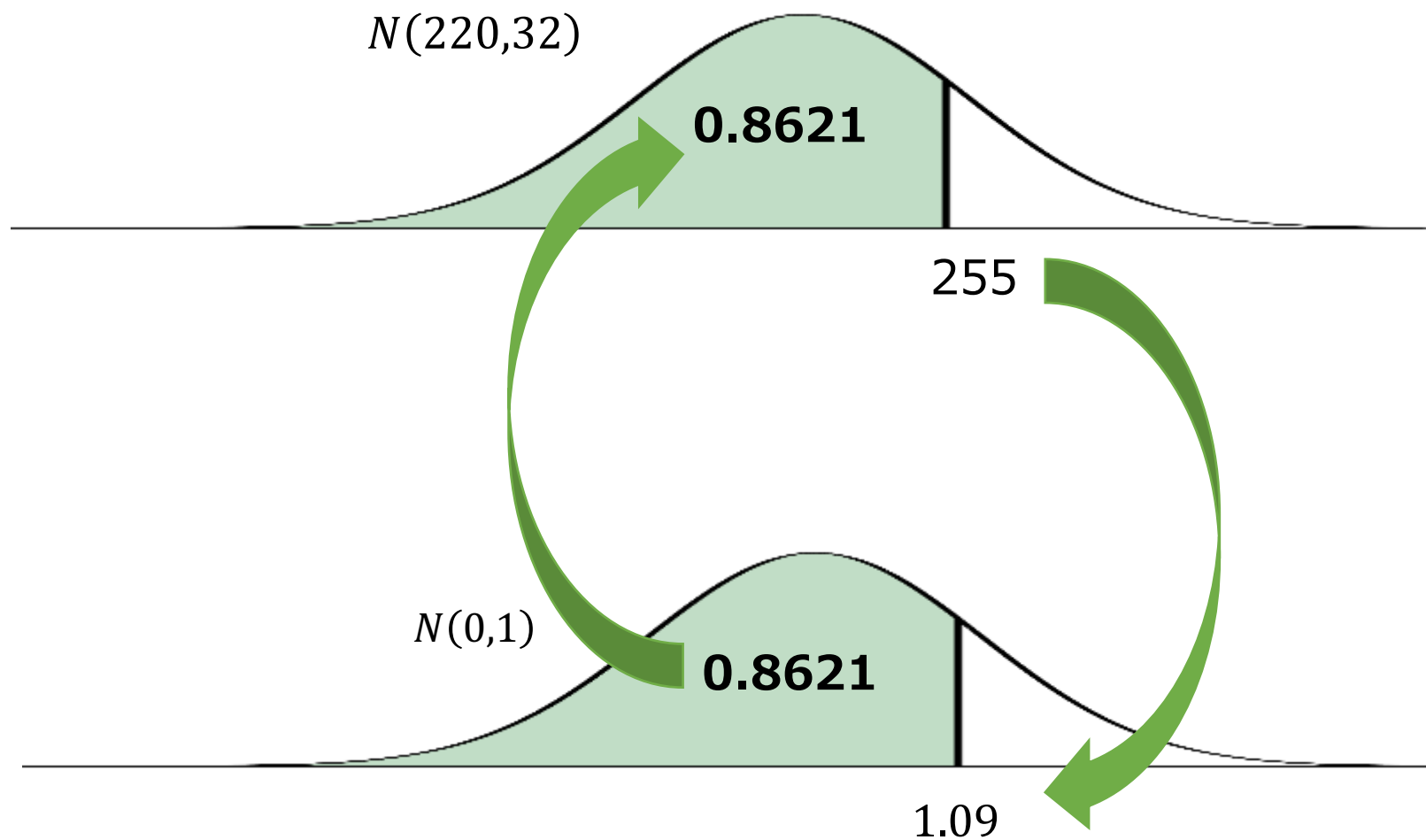
1.09

Z表の読み方



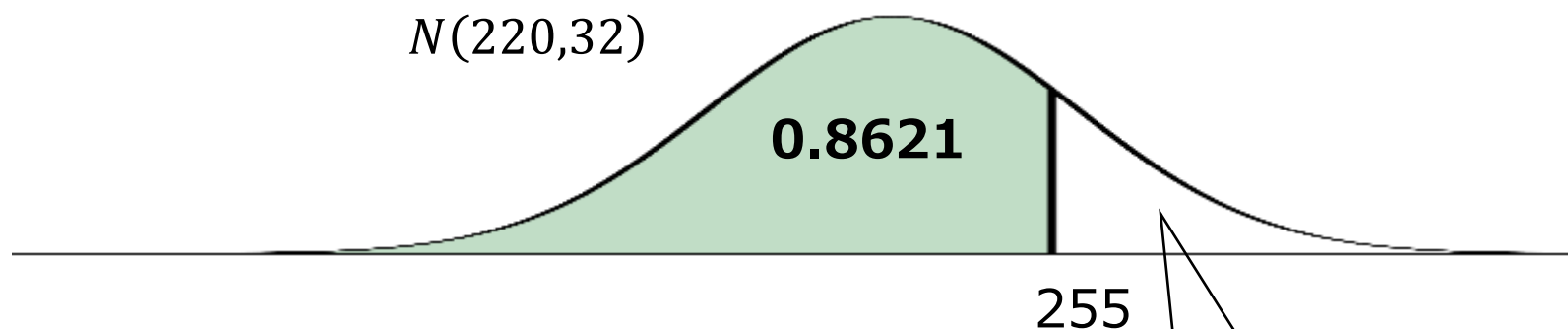
Z表を使って確率を求める

$N(220,32)$ の正規分布において、255以下の確率を求めよ



Z表を使って確率を求める

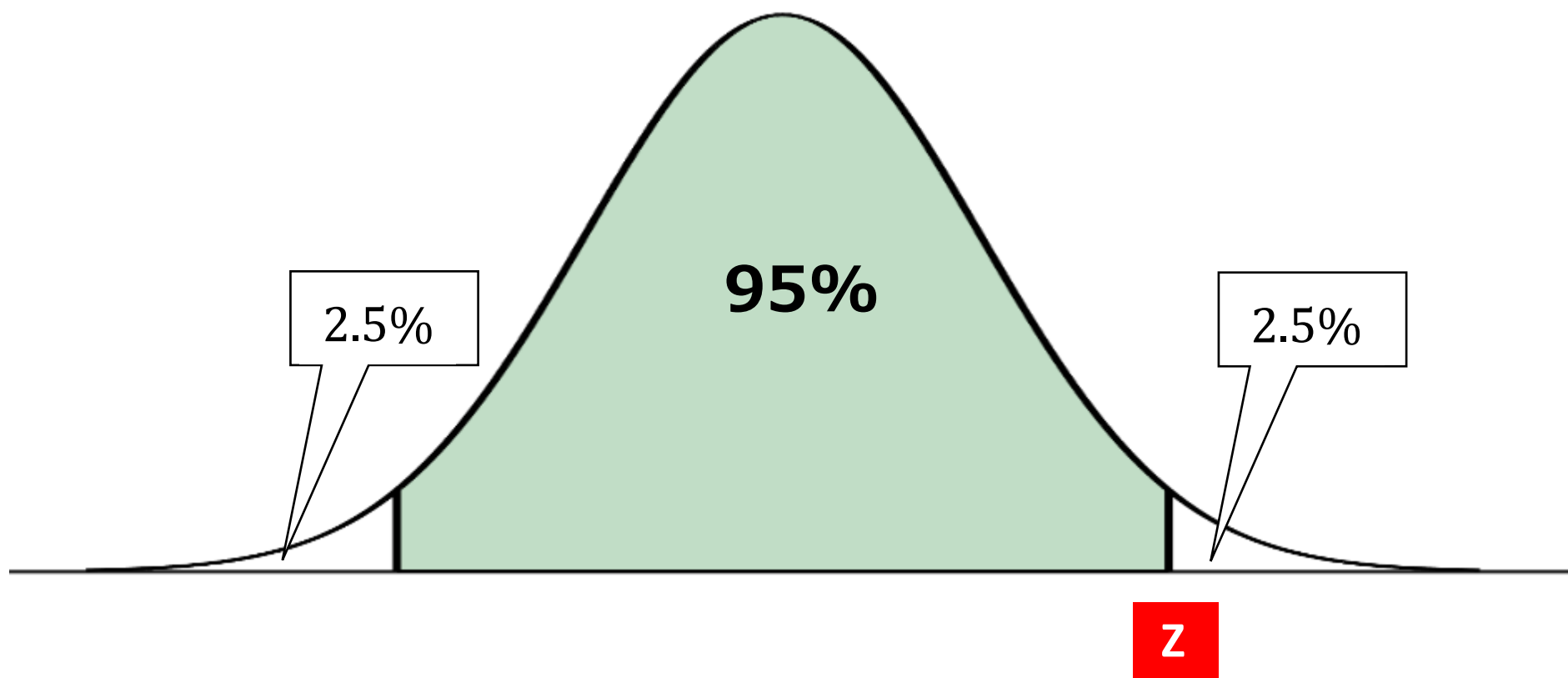
$N(220,32)$ の正規分布において、255以上の確率を求めよ



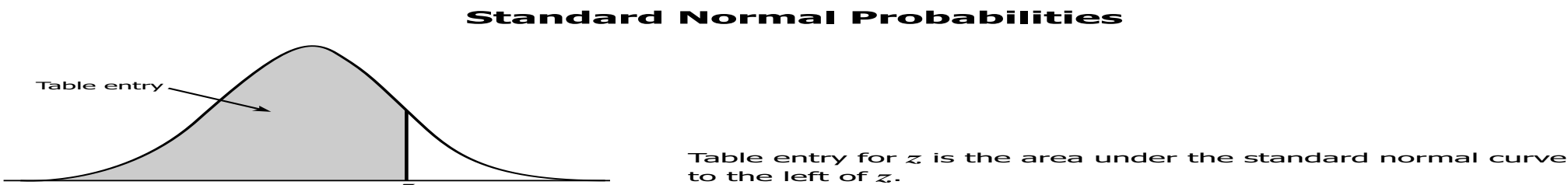
$$P(A^c) = 1 - P(A)$$

$$1 - 0.8621 = 0.1379$$

問題：zの値を特定せよ

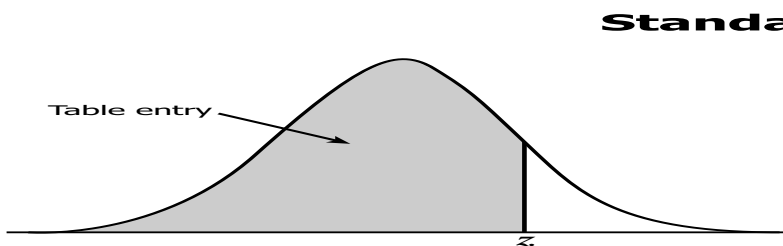


Z表の読み方



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Z表の読み方



Standard Normal Probabilities

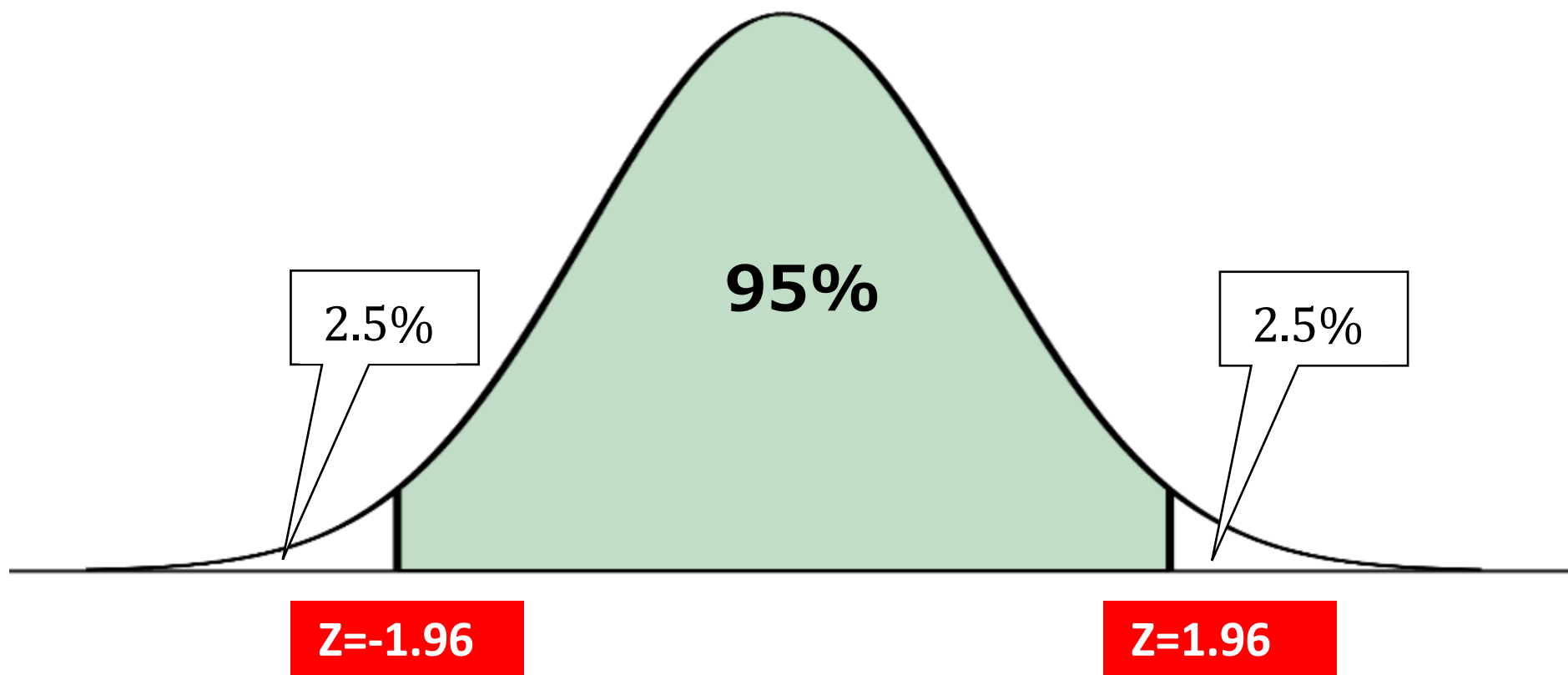
0.975を探す！

Z= 1.96

Table entry for z is the area under the standard normal curve to the left of z .

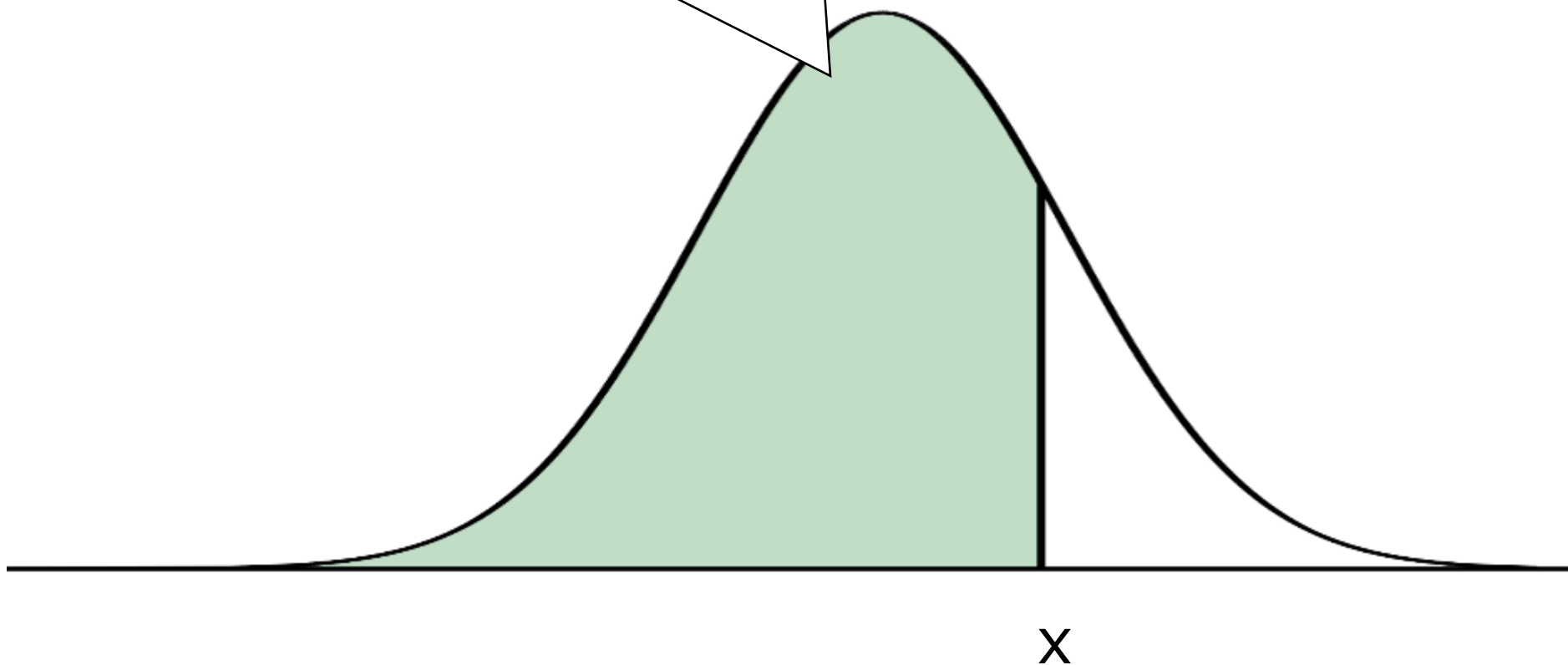
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9804	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$z=1.96$ は重要な数値！



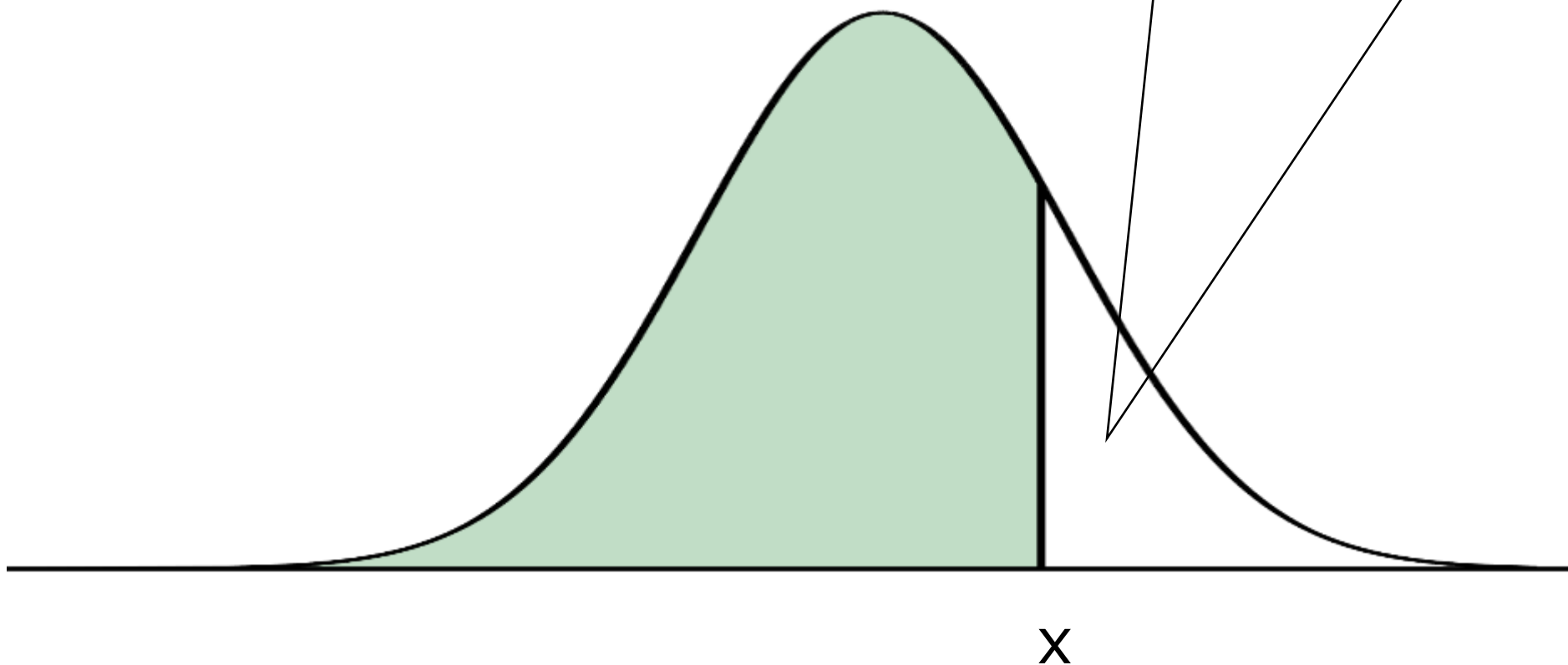
エクセル関数を使った確率の求め方

=NORM.DIST(x、平均値、標準偏差、TRUE)



エクセル関数を使った確率の求め方

=1-NORM.DIST(x、平均値、標準偏差、TRUE)



問題 1 (Excel) 演習

株価Aは平均株価が290円、標準偏差が30円の正規分布に従い、変動することがわかっています。この時、

- (1) 株価Aが280円以下を変動する確率を求めよ。
- (2) 株価Aが270円以上を変動する確率を求めよ。

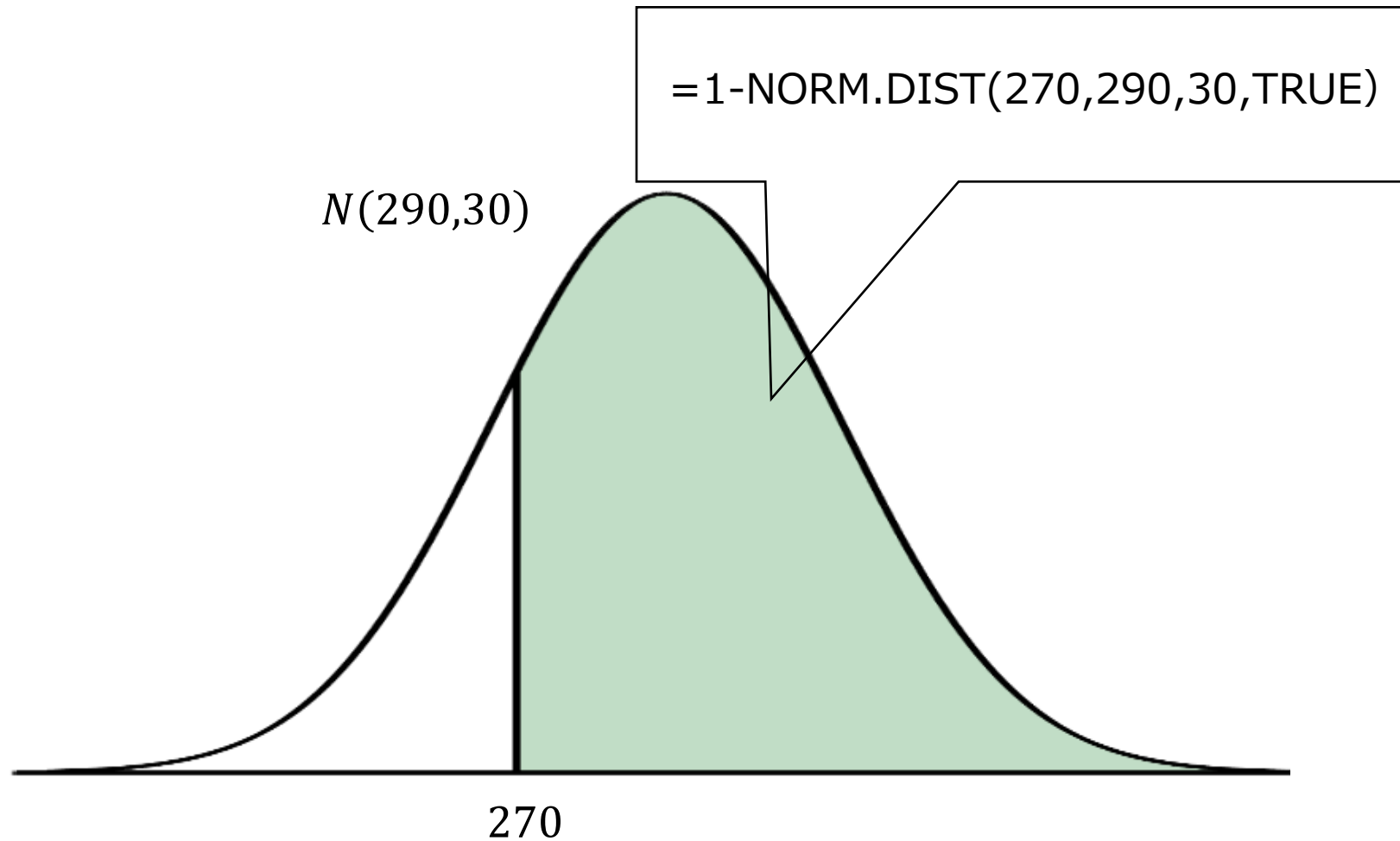
株価Aが280円以下を変動する確率

=NORM.DIST(280,290,30,TRUE)

$N(290,30)$

280

株価Aが270円以上を変動する確率

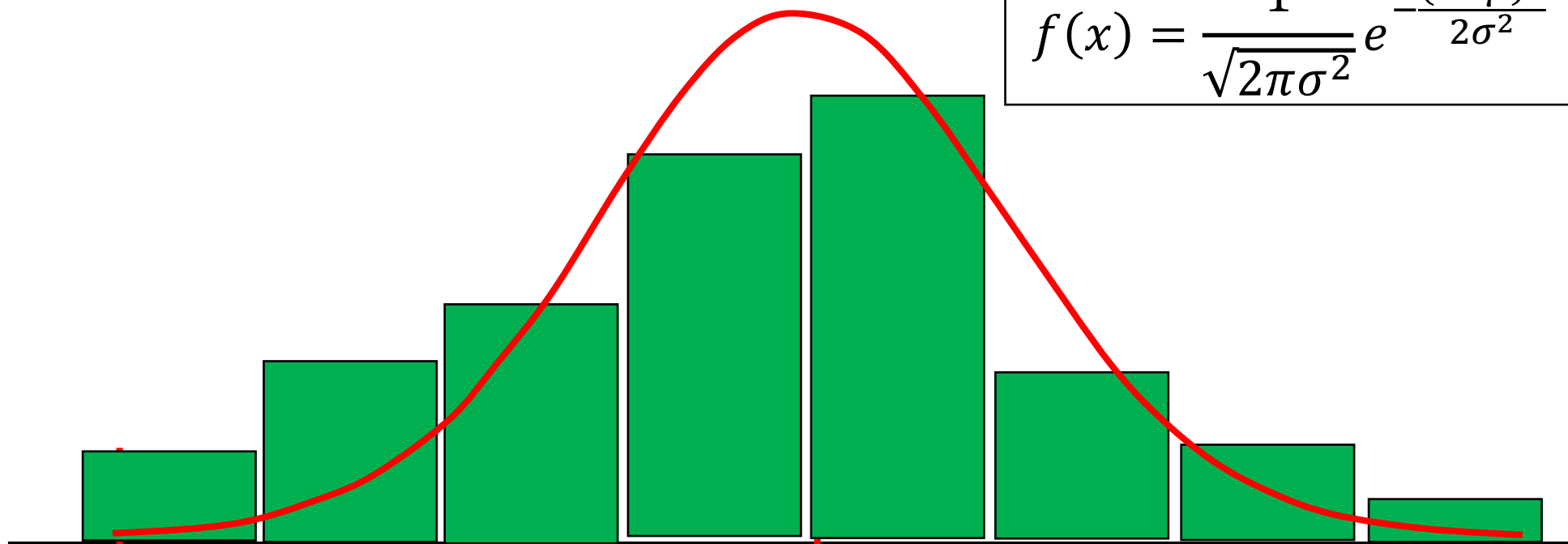


いろいろな分布

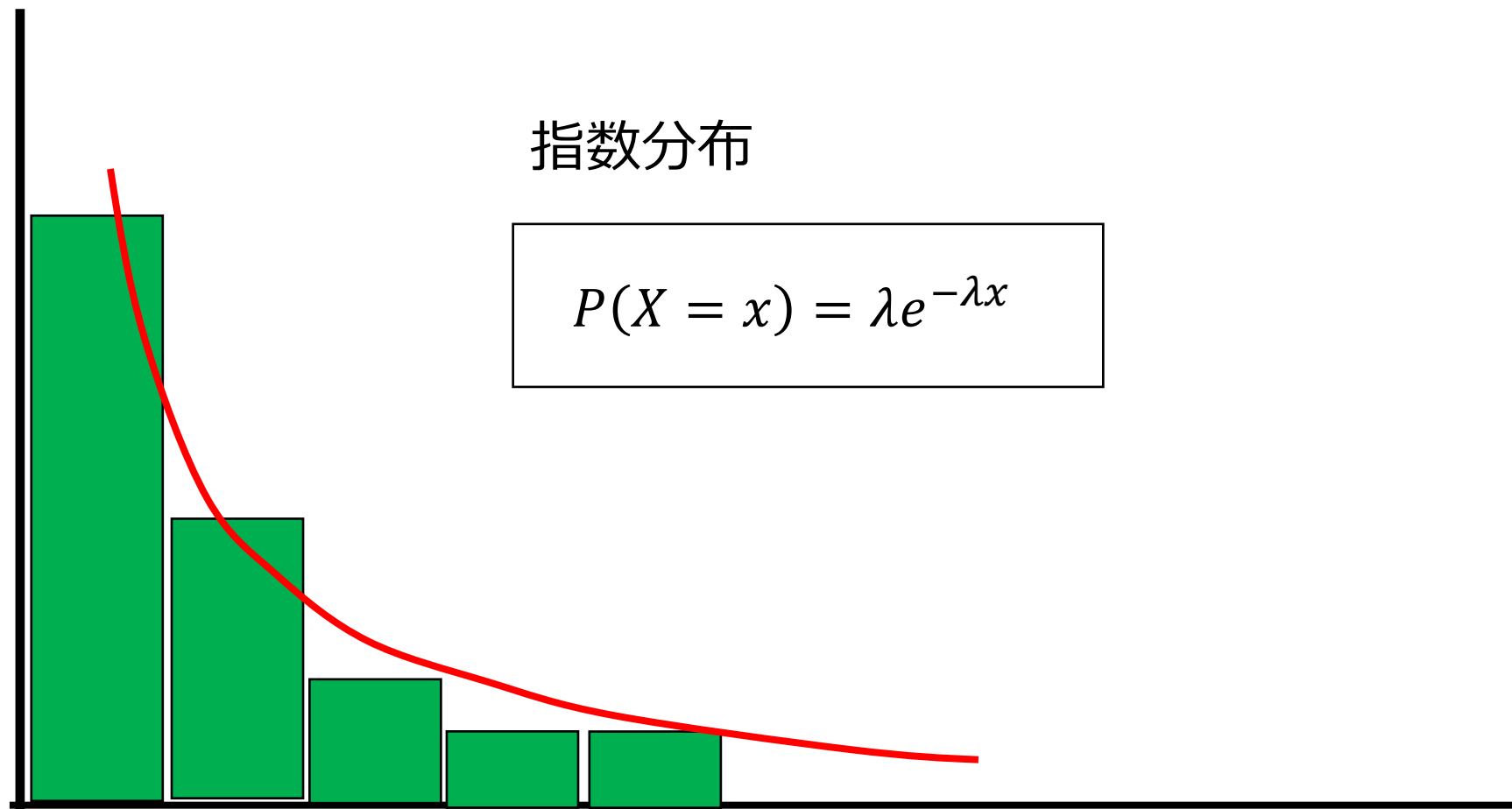
いろいろな確率分布

正規分布

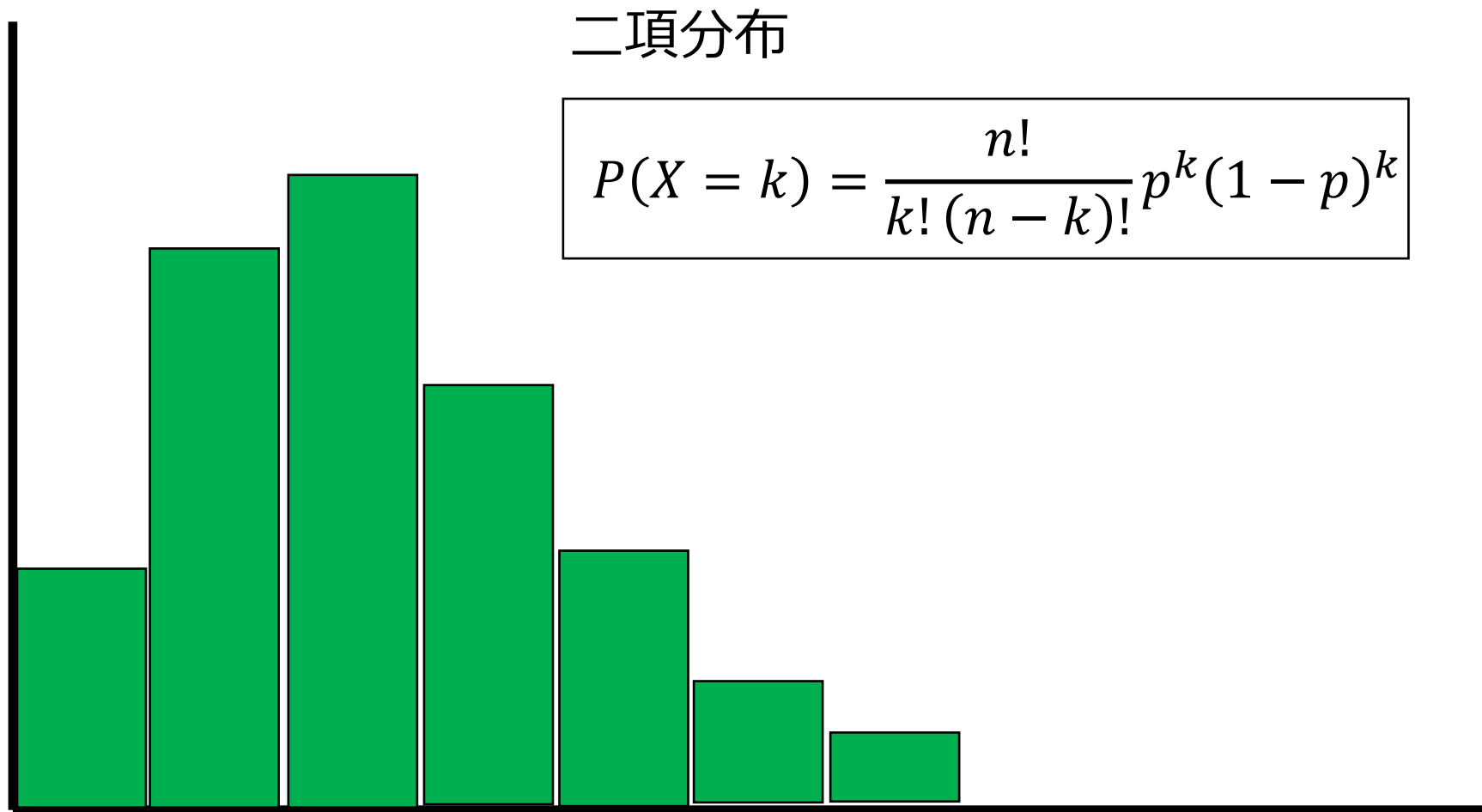
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



いろいろな確率分布



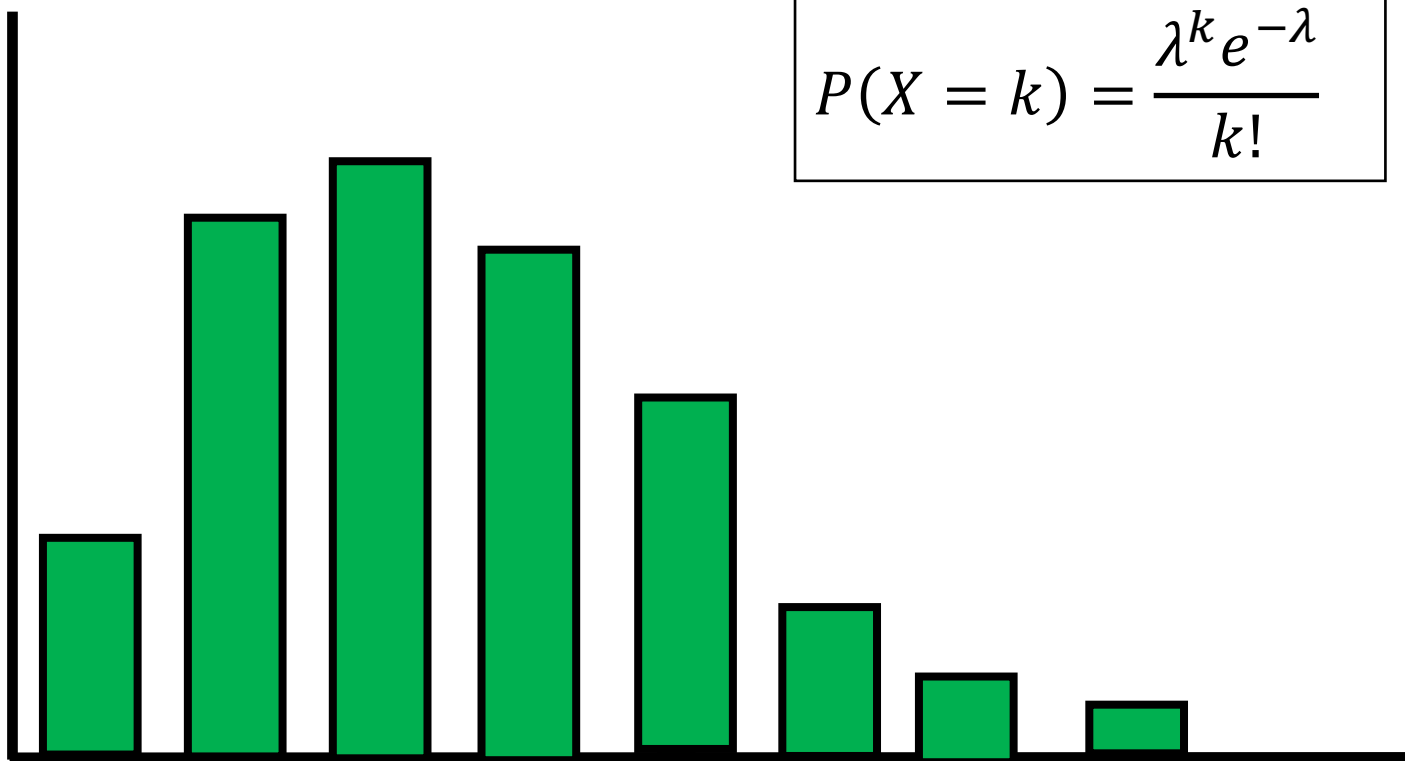
いろいろな確率分布



いろいろな確率分布

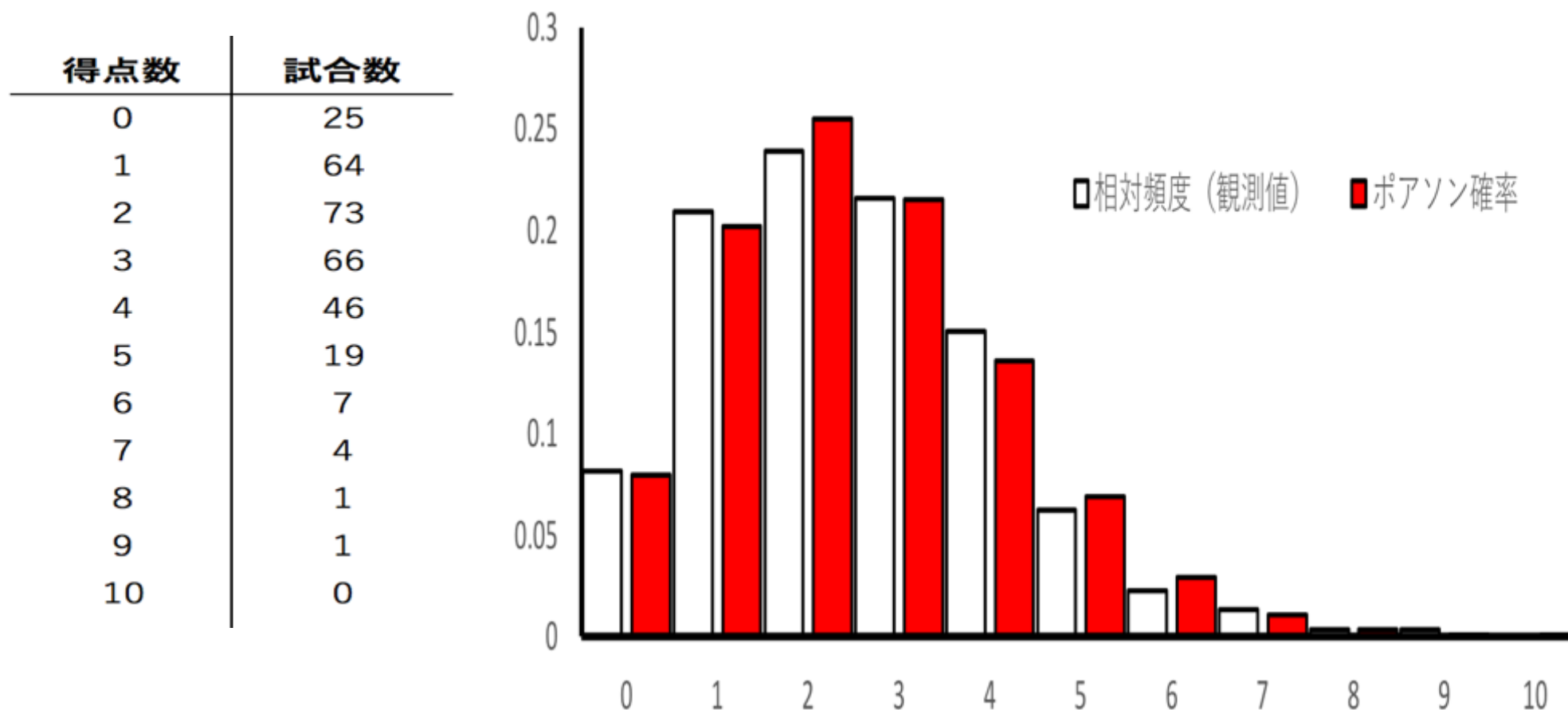
ポアソン分布

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$



ポアソン分布

下の表は2014年のJ1リーグの1試合あたりの得点数と試合数のデータです。1試合あたりの平均得点数は2.53点でした。



異常値の検出

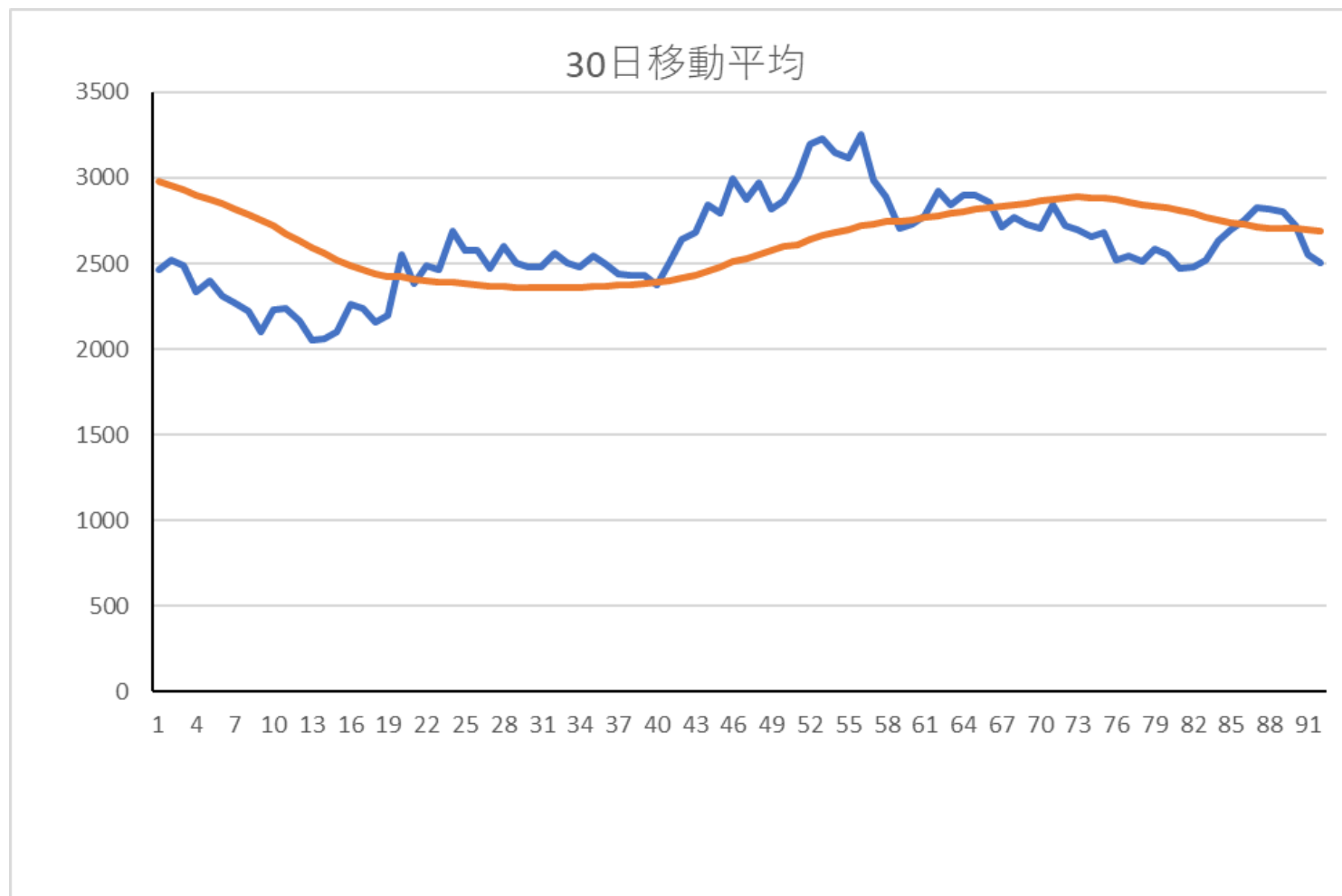
(前回) 移動平均法を使ったトレンドの抽出

課題：「アクセス数のトレンドを推定せよ」



(前回) 移動平均法を使ったトレンドの抽出

- 時系列データ = **トレンド** + 周期変動 + 不規則変動



異常値の検出

