

統計カクイズ:

そのデータから何が読みとれるのか?

(著者: 涌井良幸 発行: 実務教育出版)

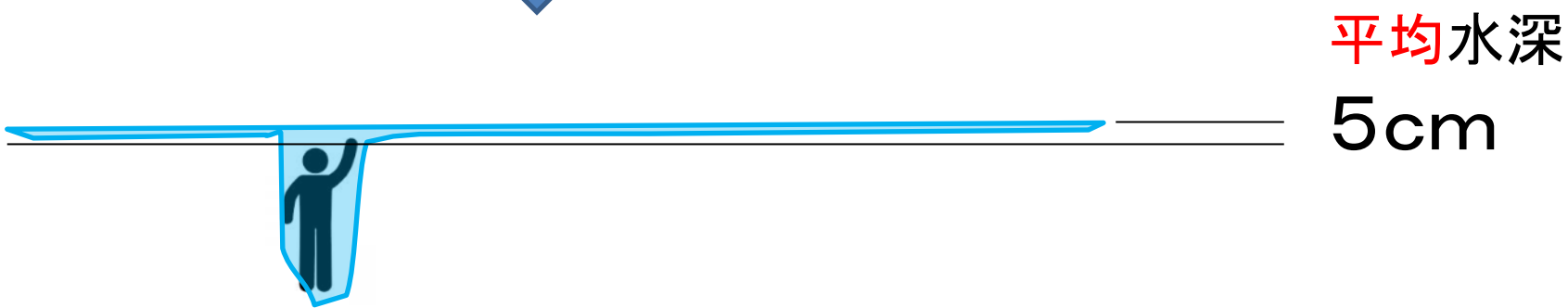
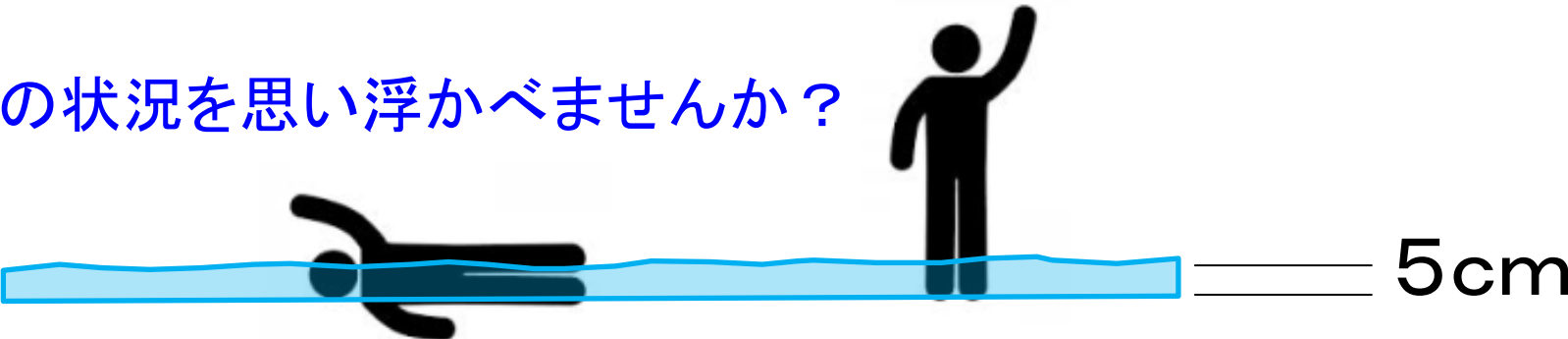
問題1

「大人が平均水深がたった5cmの池で溺れた！」というニュースは本当？ その理由を統計的に説明してください。

問題1

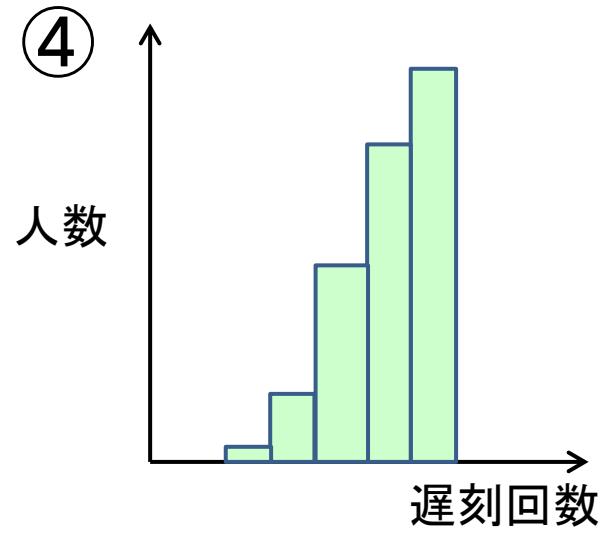
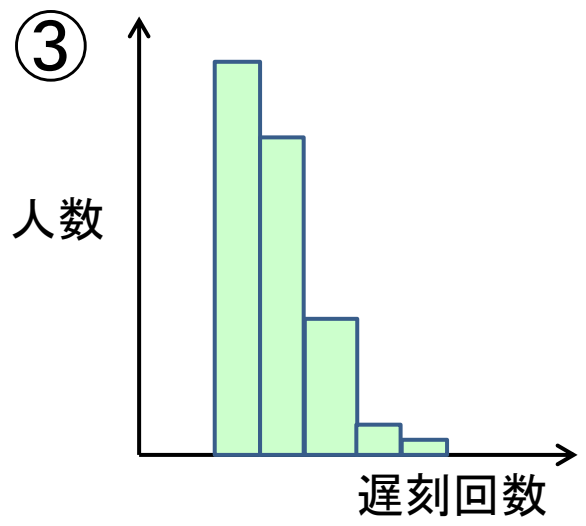
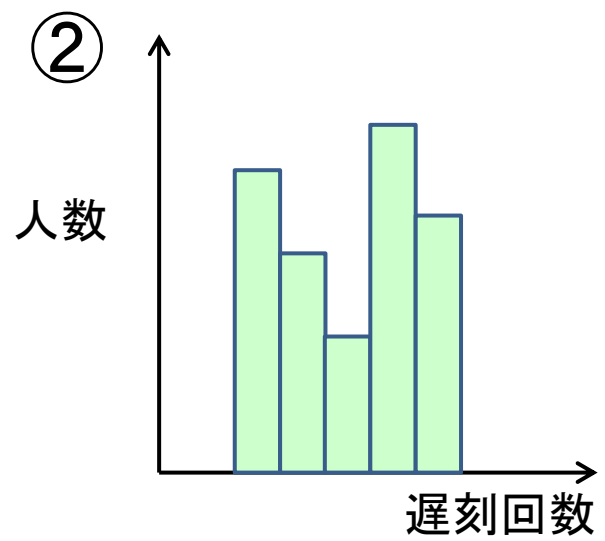
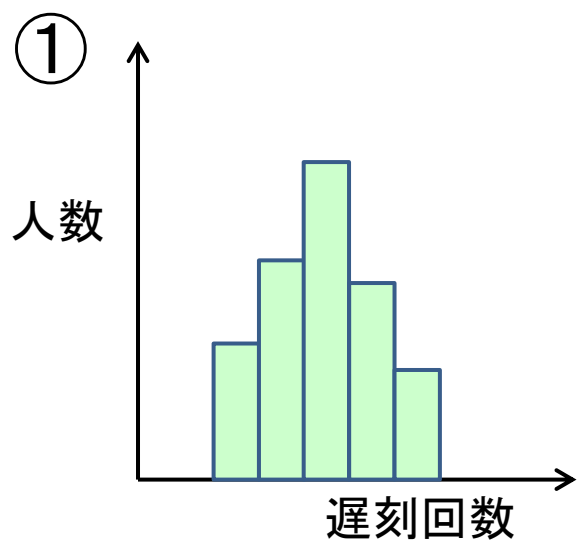
「大人が平均水深がたった5cmの池で溺れた！」というニュースは本当？ その理由を統計的に説明してください。

この状況を思い浮かべませんか？



問題2

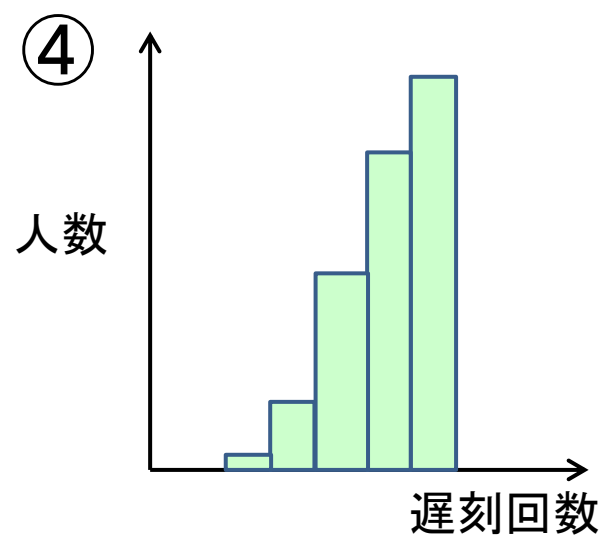
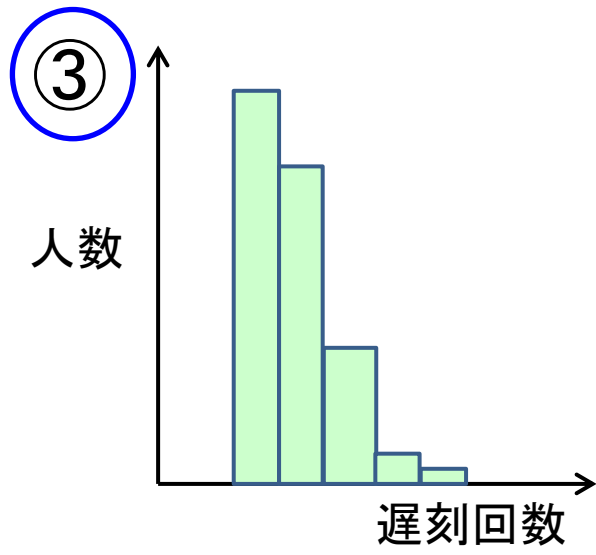
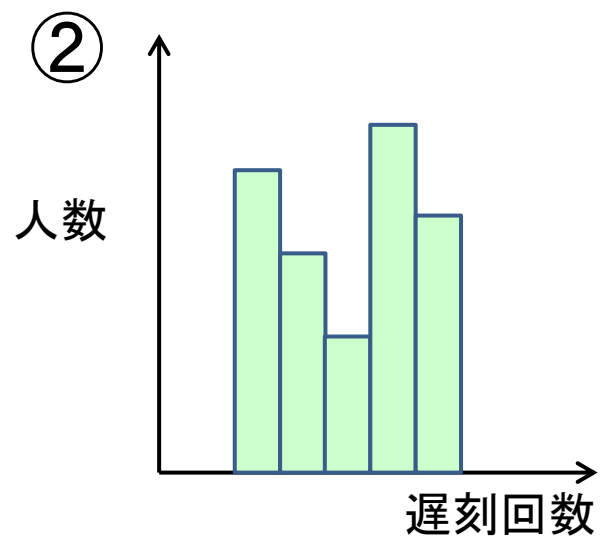
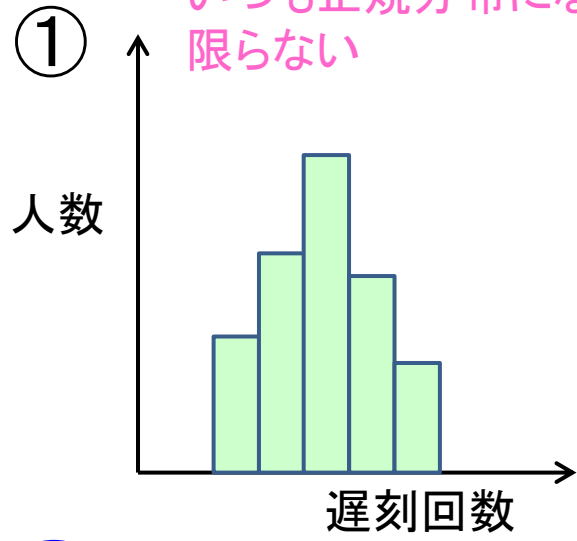
社員の遅刻回数をグラフにするとどんな形になるか？以下の4つから選りなさい。



問題2

社員の遅刻回数をグラフにするとどんな形になるか？以下の4つから選べなさい。

いつも正規分布になるとは限らない



問題3

コインの表は $1/2$ 、サイコロの目は $1/6$ 、では画鋏の針が上を向く確率は？

- ①上を向くか下を向くかの2つに1つ。よって確率は $1/2$
- ②針の長さ L 、円盤の直径 D の長さを調べる。求める確率は $L \div (L+D)$
- ③何回か投げてみて、針が上を向く回数を調べる。

問題3

コインの表は $1/2$ 、サイコロの目は $1/6$ 、では画鋲の針が上を向く確率は？

- ①上を向くか下を向くかの2つに1つ。よって確率は $1/2$
- ②針の長さ L 、円盤の直径 D の長さを調べる。求める確率は $L \div (L+D)$
- ③何回か投げてみて、針が上を向く回数を調べる。

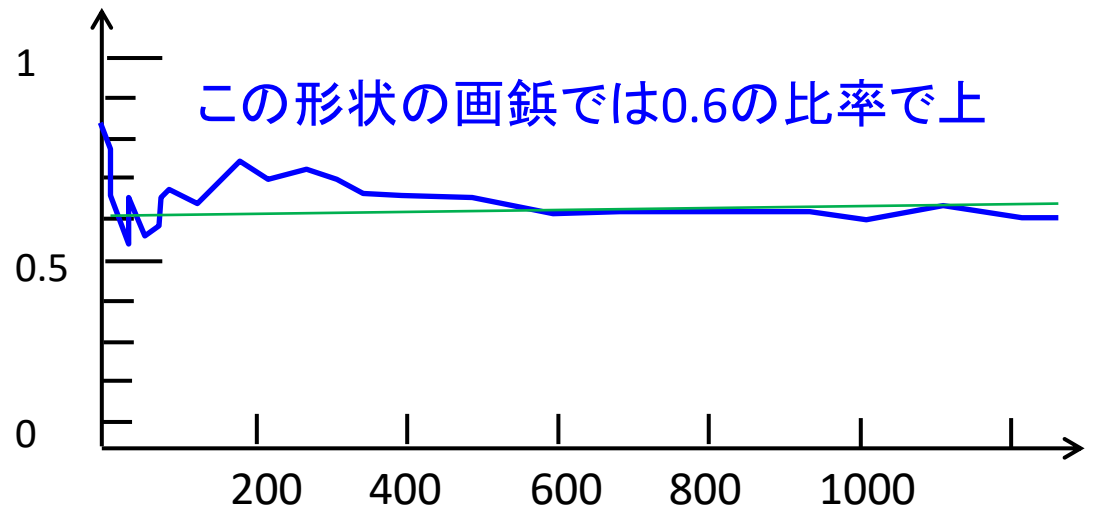


針が上

針が下

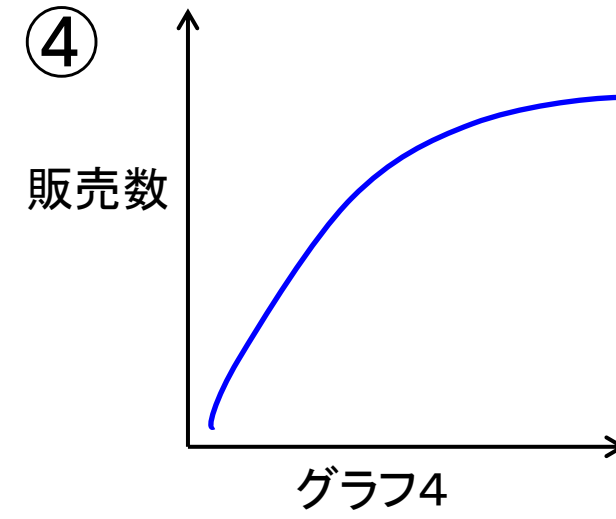
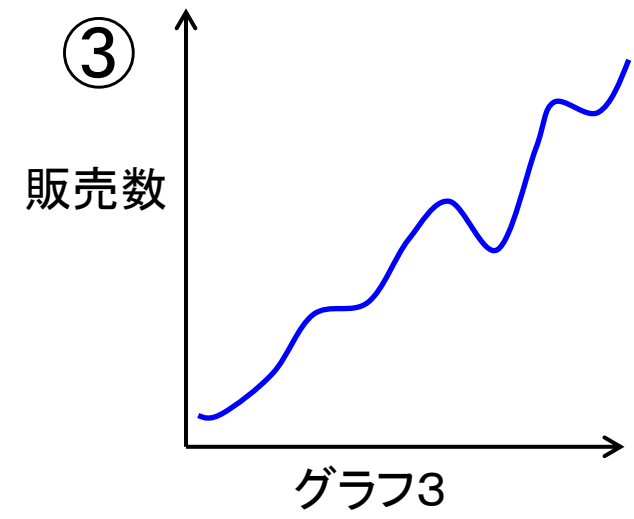
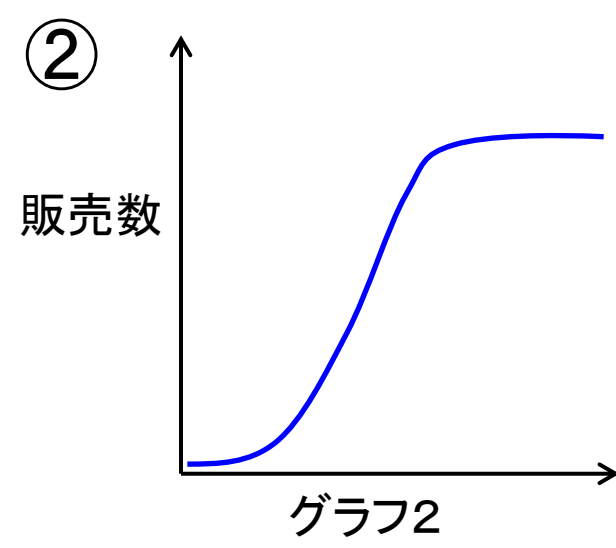
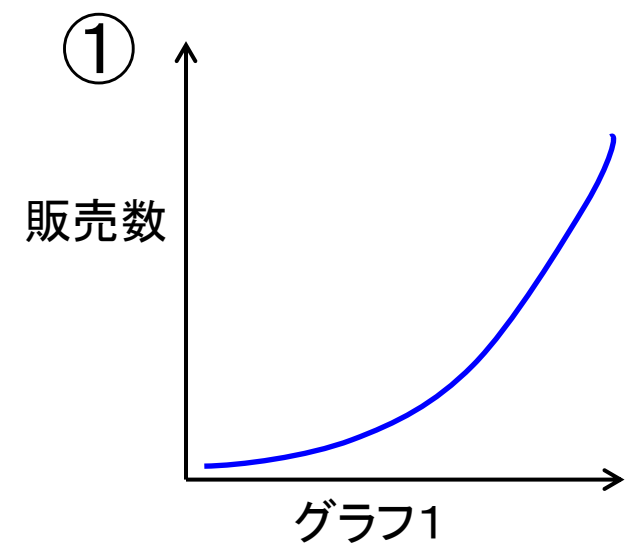
「数学的確率」: コインの裏表のように、各々が同様に確からしいとすると表あるいは裏が出る**理論的**確率は $1/2$

「統計的確率」: **実験で確かめる**



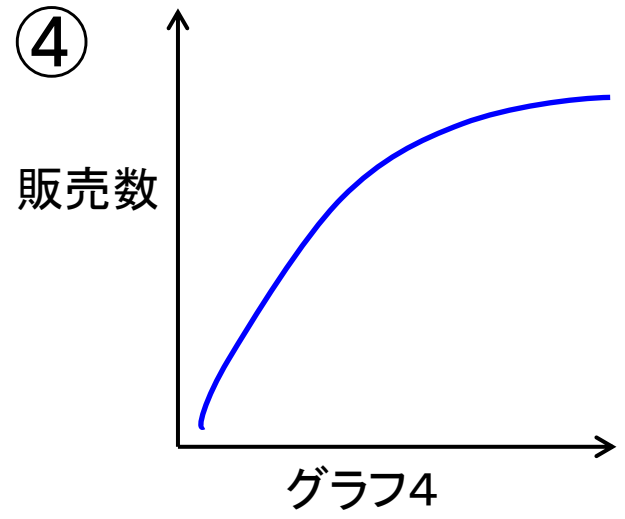
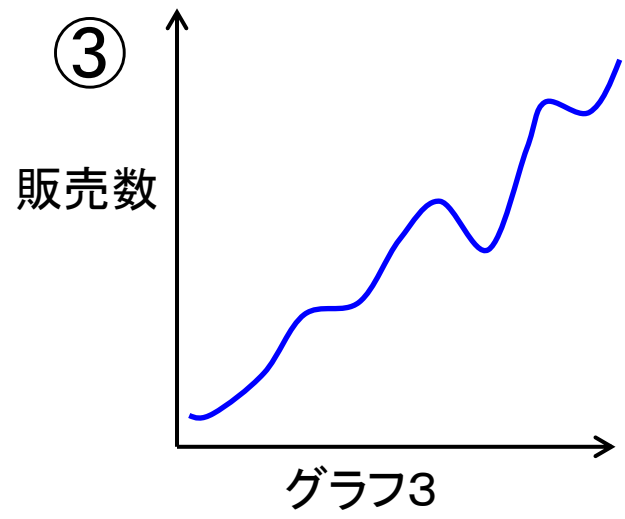
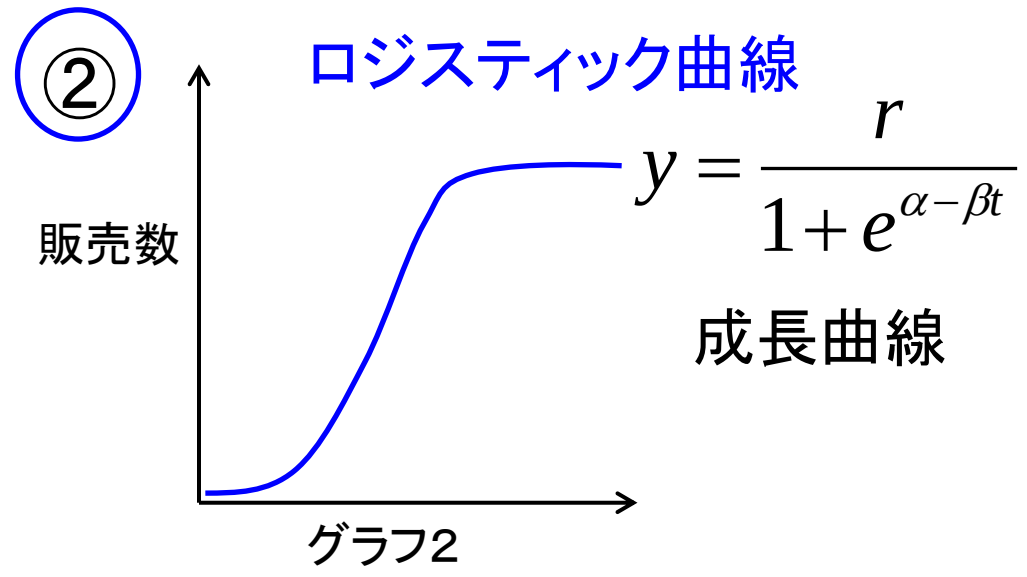
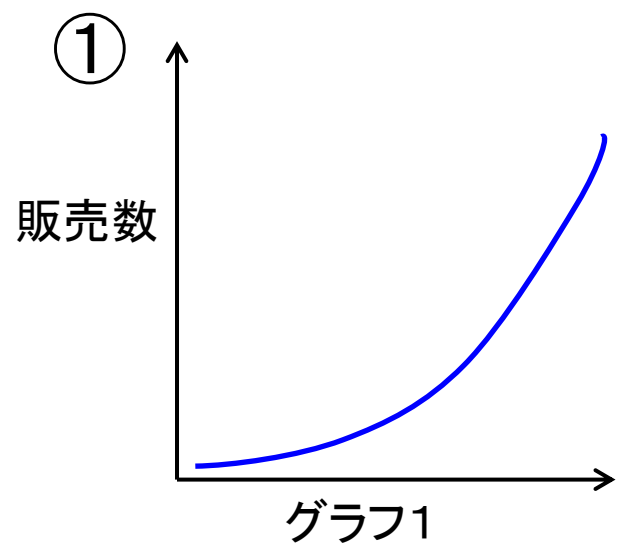
問題4

売れた商品が辿る販売曲線はどれか？



問題4

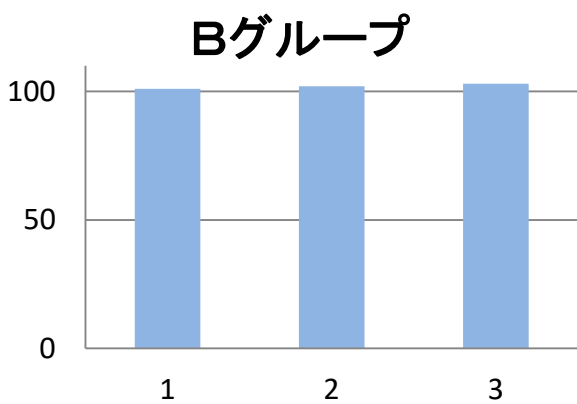
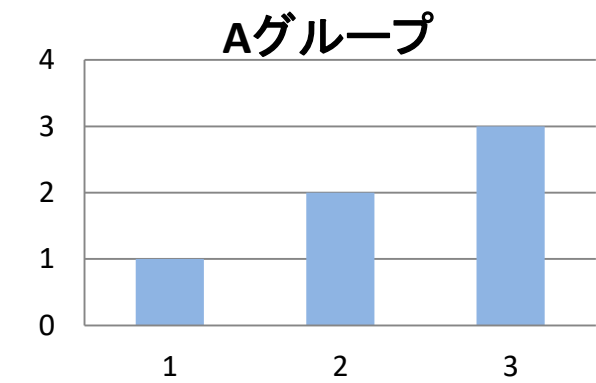
売れた商品が辿る販売曲線はどれか？



問題5

Aグループが{1, 2, 3}、Bグループが{101,102,103}
どちらのグループがバラツキが大きい？

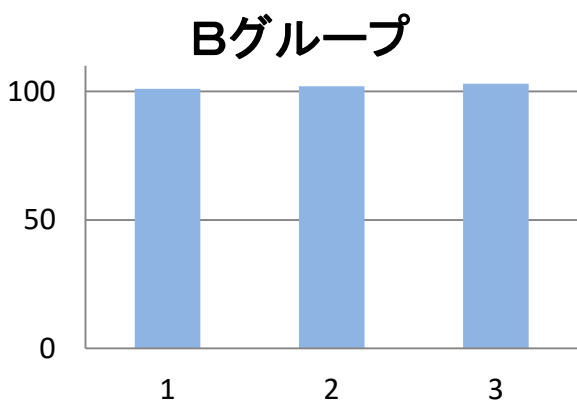
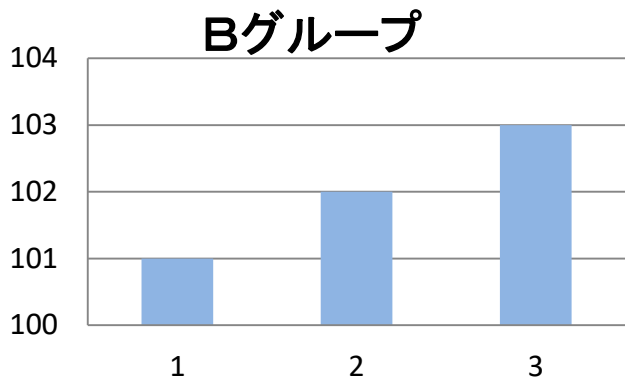
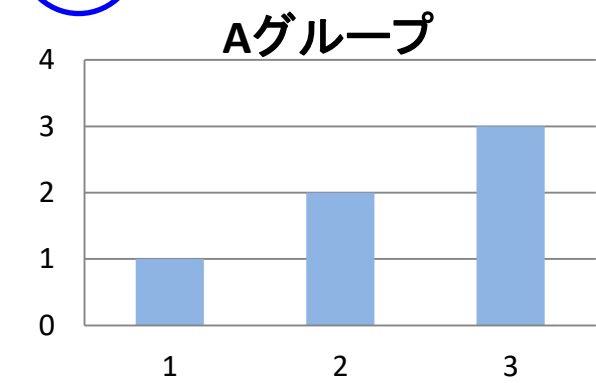
- ①Aグループが大きい
- ②Bグループが大きい
- ③同じ



問題5

Aグループが{1, 2, 3}、Bグループが{101,102,103}
どちらのグループがバラツキが大きい？

- ①Aグループが大きい
- ②Bグループが大きい
- ③同じ



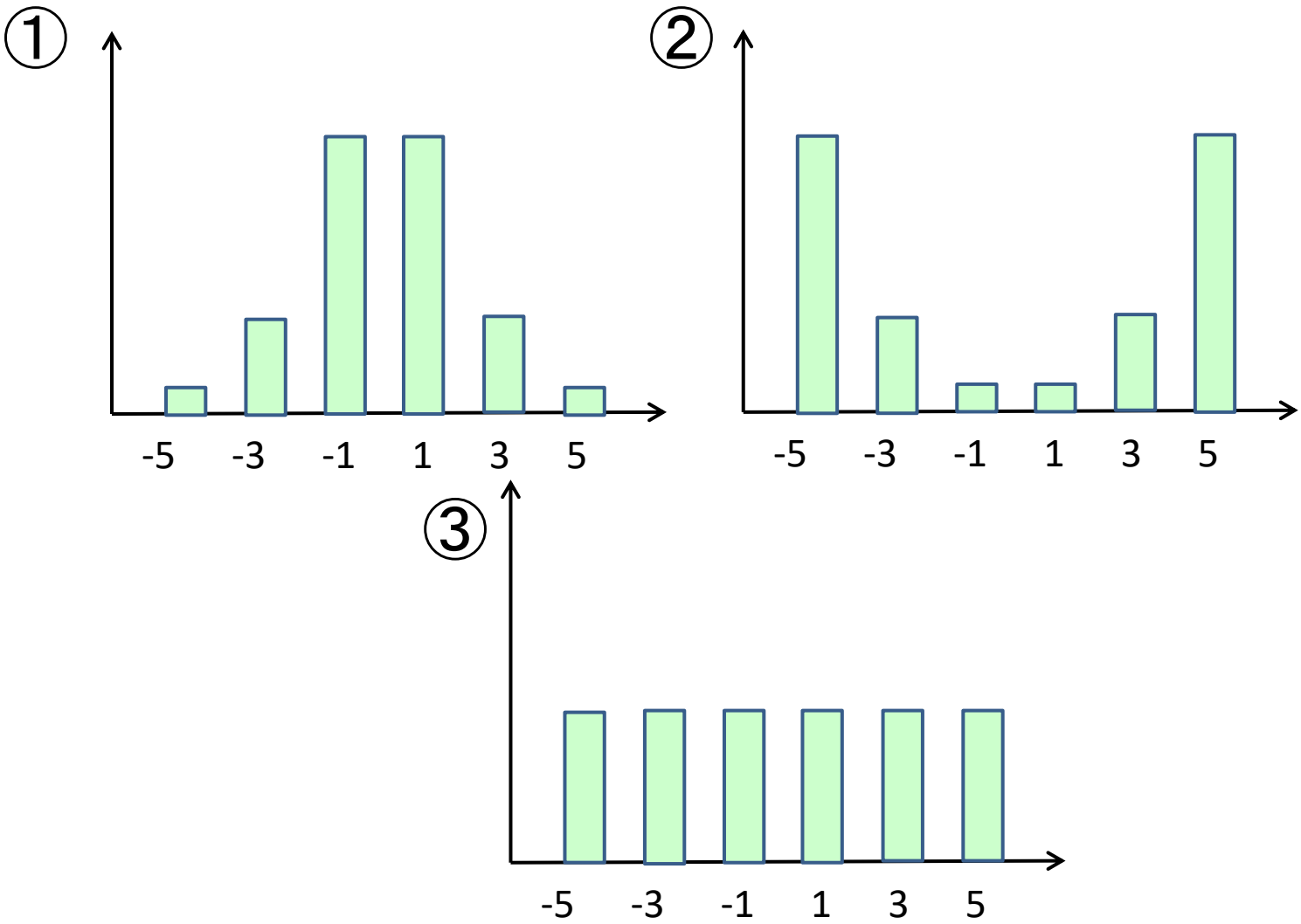
分散 $V = \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$

分散 $V_A = \sigma_A^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = \frac{2}{3} = 0.66\dots$

分散 $V_B = \sigma_B^2 = \frac{(101-102)^2 + (102-102)^2 + (103-102)^2}{3} = \frac{2}{3} = 0.66\dots$

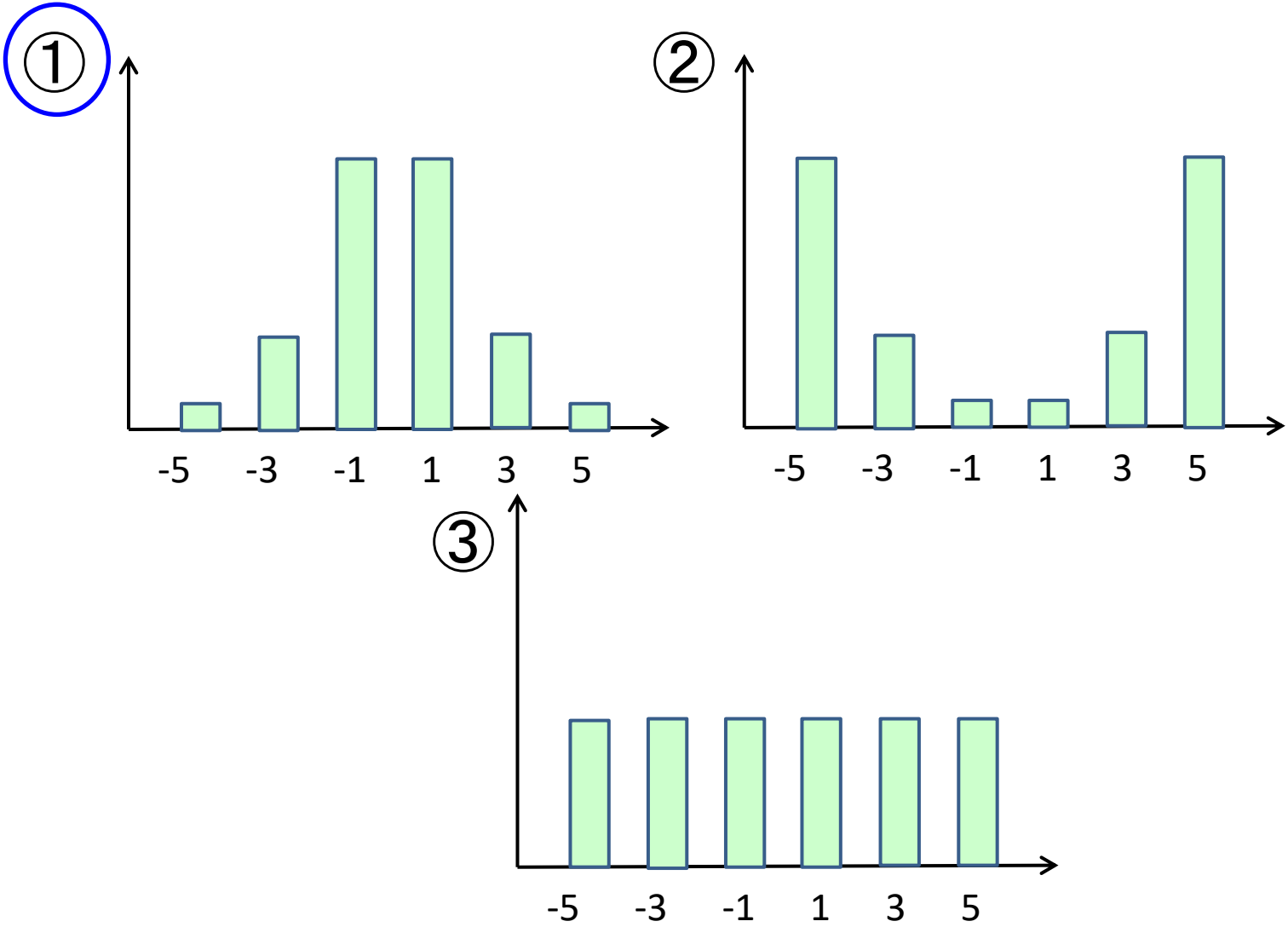
問題6

酔っ払いのランダムウォークの行き先はどこ？
最初は原点、確率0.5で右、確率0.5で左に1歩ずつ移動を繰り返します。
5回移動後、この酔っ払いはどこに到達するか？ グラフを選んでください。



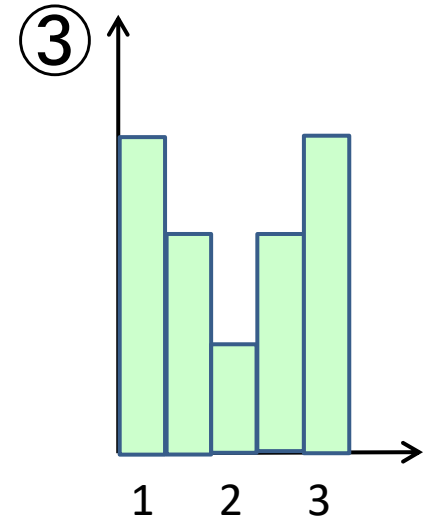
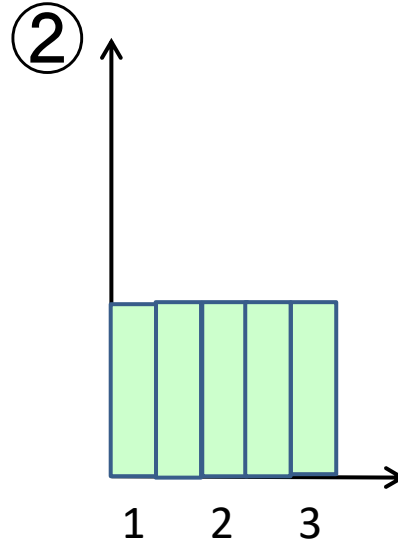
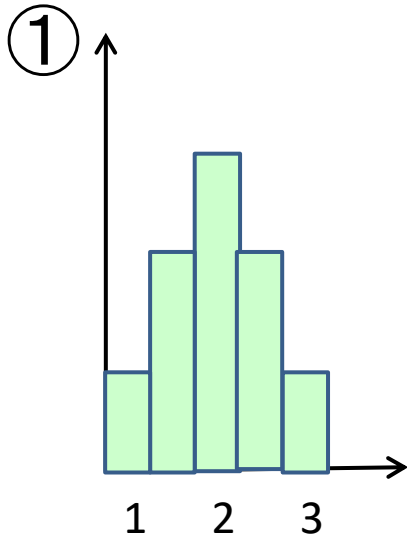
問題6

酔っ払いのランダムウォークの行き先はどこ？
最初は原点、確率0.5で右、確率0.5で左に1歩ずつ移動を繰り返します。
5回移動後、この酔っ払いはどこに到達するか？ グラフを選んでください。



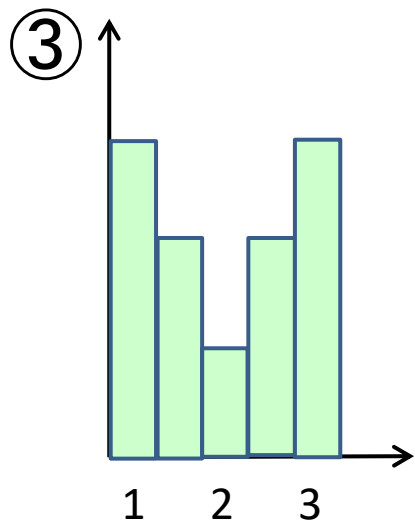
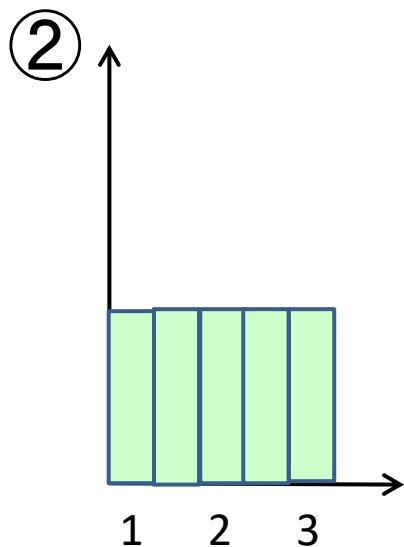
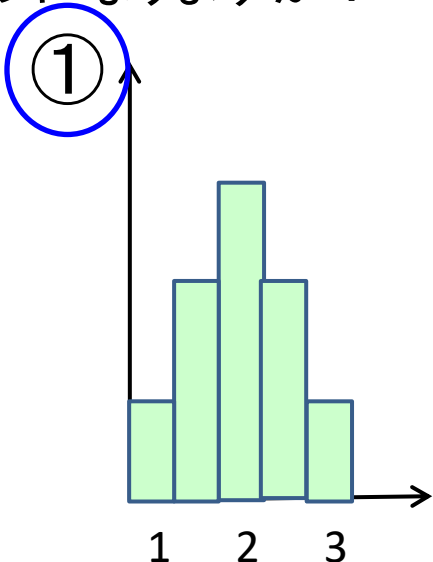
問題7

1、2、3と書かれた3種類のカードが1枚ずつあります。1枚引いて、元に戻し、もう一回引く。この平均値を算出します。何回か繰り返したあとの度数分布はどのグラフになりますか？



問題7

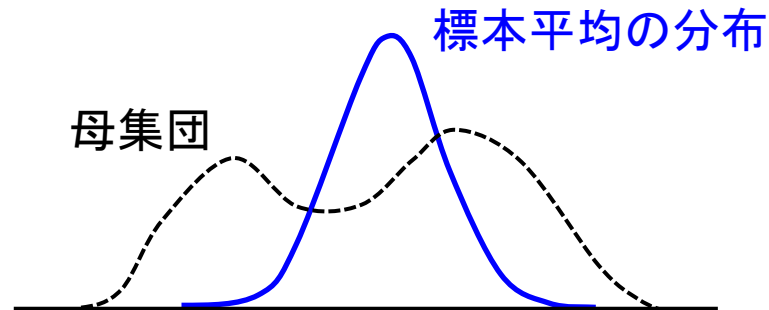
1、2、3と書かれた3種類のカードが1枚ずつあります。1枚引いて、元に戻し、もう一回引く。この平均値を算出します。何回か繰り返したあとの度数分布はどのグラフになりますか？



(1, 1)(1, 2)(1, 3)(2, 1)(2, 2)(2, 3)(3, 1)(3, 2)(3, 3)の9通り

$\bar{X}$ の値	$\frac{2}{2}$	$\frac{3}{2}$	$\frac{4}{2}$	$\frac{5}{2}$	$\frac{6}{2}$
$\bar{X}$ の度数	1	2	3	2	1
$\bar{X}$ の確率	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

「中心極限定理」母集団から取り出された標本の標本平均の分布は母集団がどんな分布でもほぼ正規分布になる



問題8

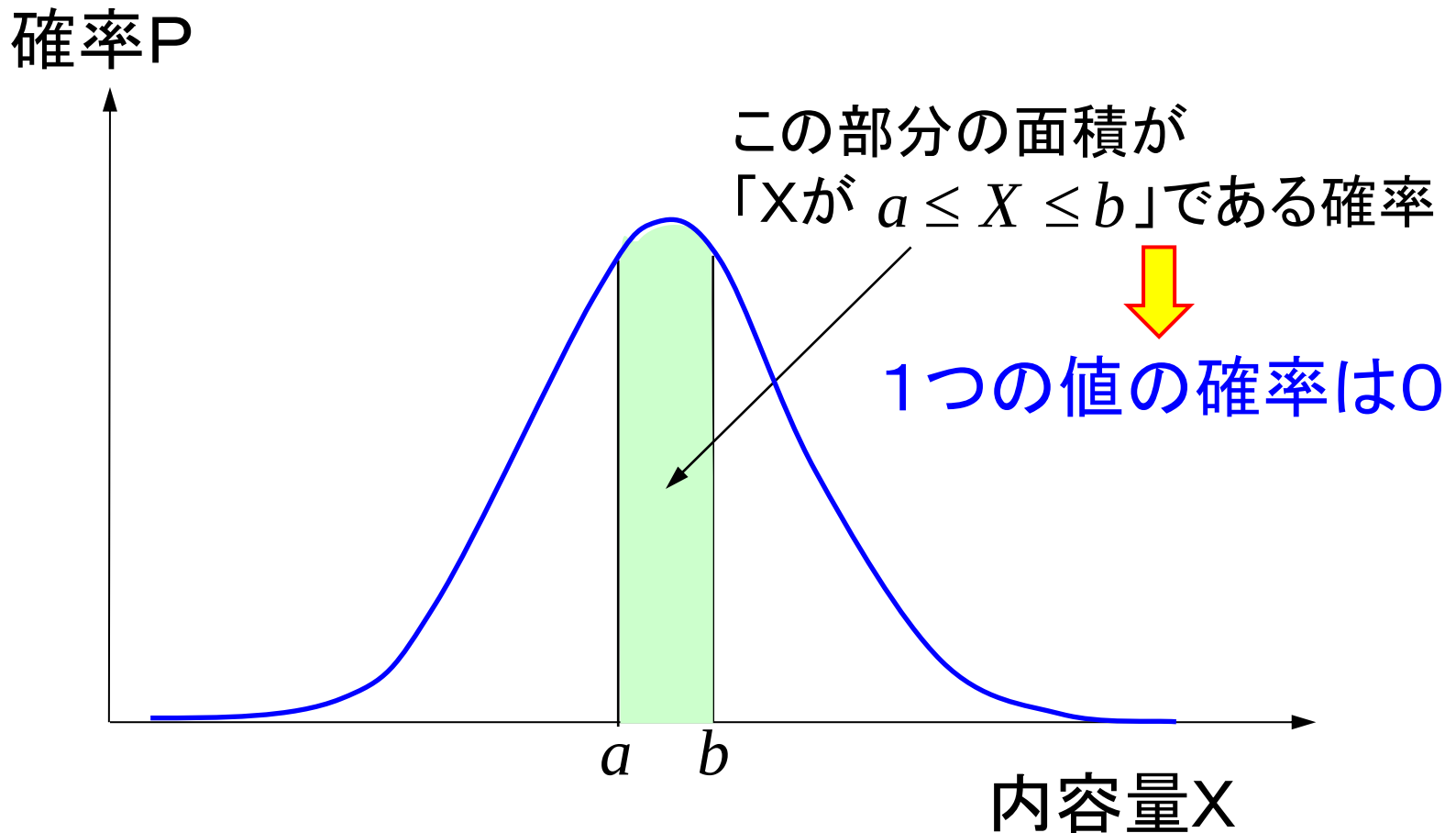
下図は200mLのペットボトルに注入された内容量の分布である。このペットボトルにちょうど200mL入っている確率はどのくらいか？

- ①確率は、グラフのように一番大きく 0.5
- ②確率は、0でない何らかの値(大きくはない)
- ③確率は、0

問題8

下図は200mLのペットボトルに注入された内容量の分布である。このペットボトルにちょうど200mL入っている確率はどのくらいか？

- ① 確率は、グラフのように一番大きく 0.5
- ② 確率は、0でない何らかの値(大きくはない)
- ③ 確率は、0



問題9

偏差値とは、その平均値が50、標準偏差が10である。

偏差値50は並、60は上位、70は秀才、80は天才と言われる。

100を超える偏差値や、マイナスの偏差値はあり得るか？

①100超え、0未満の偏差値は「ある」

①100超え、0未満の偏差値は「ない」

問題9

偏差値とは、その平均値が50、標準偏差が10である。

偏差値50は並、60は上位、70は秀才、80は天才と言われる。

100を超える偏差値や、マイナスの偏差値はあり得るか？

① 100超え、0未満の偏差値は「ある」

① 100超え、0未満の偏差値は「ない」

右表のように極端な場合はあり得る。

No.	得点	偏差値		No.	得点	偏差値
1	100	51.9		1	1	48.1
2	100	51.9		2	1	48.1
3	100	51.9		3	1	48.1
4	100	51.9		4	1	48.1
5	100	51.9		5	1	48.1
6	100	51.9		6	1	48.1
7	100	51.9		7	1	48.1
8	100	51.9		8	1	48.1
9	100	51.9		9	1	48.1
10	100	51.9		10	1	48.1
11	100	51.9		11	1	48.1
12	100	51.9		12	1	48.1
13	100	51.9		13	1	48.1
14	100	51.9		14	1	48.1
15	100	51.9		15	1	48.1
16	100	51.9		16	1	48.1
17	100	51.9		17	1	48.1
18	100	51.9		18	1	48.1
19	100	51.9		19	1	48.1
20	100	51.9		20	1	48.1
21	100	51.9		21	1	48.1
22	100	51.9		22	1	48.1
23	100	51.9		23	1	48.1
24	100	51.9		24	1	48.1
25	100	51.9		25	1	48.1
26	100	51.9		26	1	48.1
27	100	51.9		27	1	48.1
28	100	51.9		28	1	48.1
29	100	51.9		29	1	48.1
30	1	-3.9		30	100	103.9
平均	96.7	50		平均	4.3	50
標準偏差	17.77104	10		標準偏差	17.77104	10

問題10

統計学で使われる「自由度」は

- ①母集団から標本を抽出するときの標本のサイズ
- ②複数の変数が自由に变化できる度合い
- ③一度に分析できる資料の種類の上限

問題10

統計学で使われる「自由度」は

- ①母集団から標本を抽出するときの標本のサイズ
- ②複数の変数が自由に変化できる度合い
- ③一度に分析できる資料の種類の上限

3つの変数 X_1 、 X_2 及び X_3 があるとする。これらの間に何も制約がなければ、3つの変数は好き勝手に自由な値をとることができる。

この場合、自由度は3。

$X_1 + X_2 + X_3 = 5$ という条件がつくと、3つのうち2つの変数が決まると残りの1つの変数は決まってしまう。この時自由度は2。

「自由度」とは、「自由に動ける変数の個数」である。

$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ のような制約が1つあると自由度は $n-1$ となる。

$$\text{分散 } \sigma^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

←母集団の分散

$$\text{不偏分散 } \sigma^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

←標本の分散

未知なので制約されない

$$\text{分散 } \sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_n - \mu)^2}{n}$$

←母集団の分散

$$\text{不偏分散 } \hat{\sigma}^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

←標本の分散

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \quad \leftarrow \text{制約条件}$$

	標本の平均値 $\bar{x}$	母集団の平均値 μ
標本	2	2.1
1	-1	-1.1
2	0	-0.1
3	1	0.9
平方和	2.00	2.03

分散は平方和を自由度 $n-1$ で割って修正する

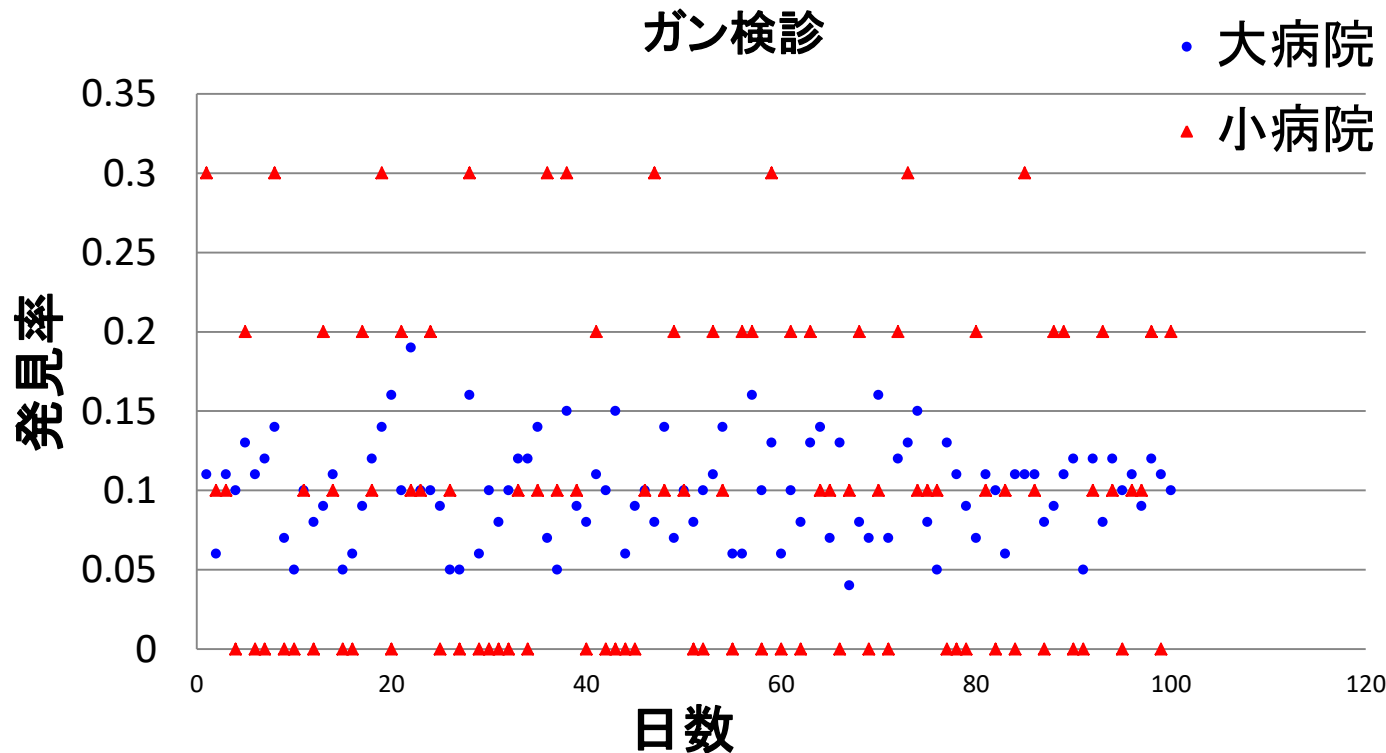
問題11

検査機器やスタッフが同レベルの大小2つの病院がある。
大病院では、ガン検診に毎日100人、小病院では10人訪れる。
10人に1人がガンであるとき、1日の検診者のうち2割以上にガン
が発見されることもあるだろう。1年で比較したとき、「検診者の2割
以上にガン発見」という日数は、大病院、小病院のどちらが多い
か？ ①大病院の方が多い
②小病院の方が多い
③大病院も小病院もほぼ同じ

問題11

検査機器やスタッフが同レベルの大小2つの病院がある。
大病院では、ガン検診に毎日100人、小病院では10人訪れる。
10人に1人がガンであるとき、1日の検診者のうち2割以上にガン
が発見されることもあるだろう。1年で比較したとき、「検診者の2割
以上にガン発見」という日数は、大病院、小病院のどちらが多い
か？

- ① 大病院の方が多い
- ② 小病院の方が多い
- ③ 大病院も小病院もほぼ同じ



小病院の方が
バラツキが大きい
ため2割を超すこと
が多い

問題12

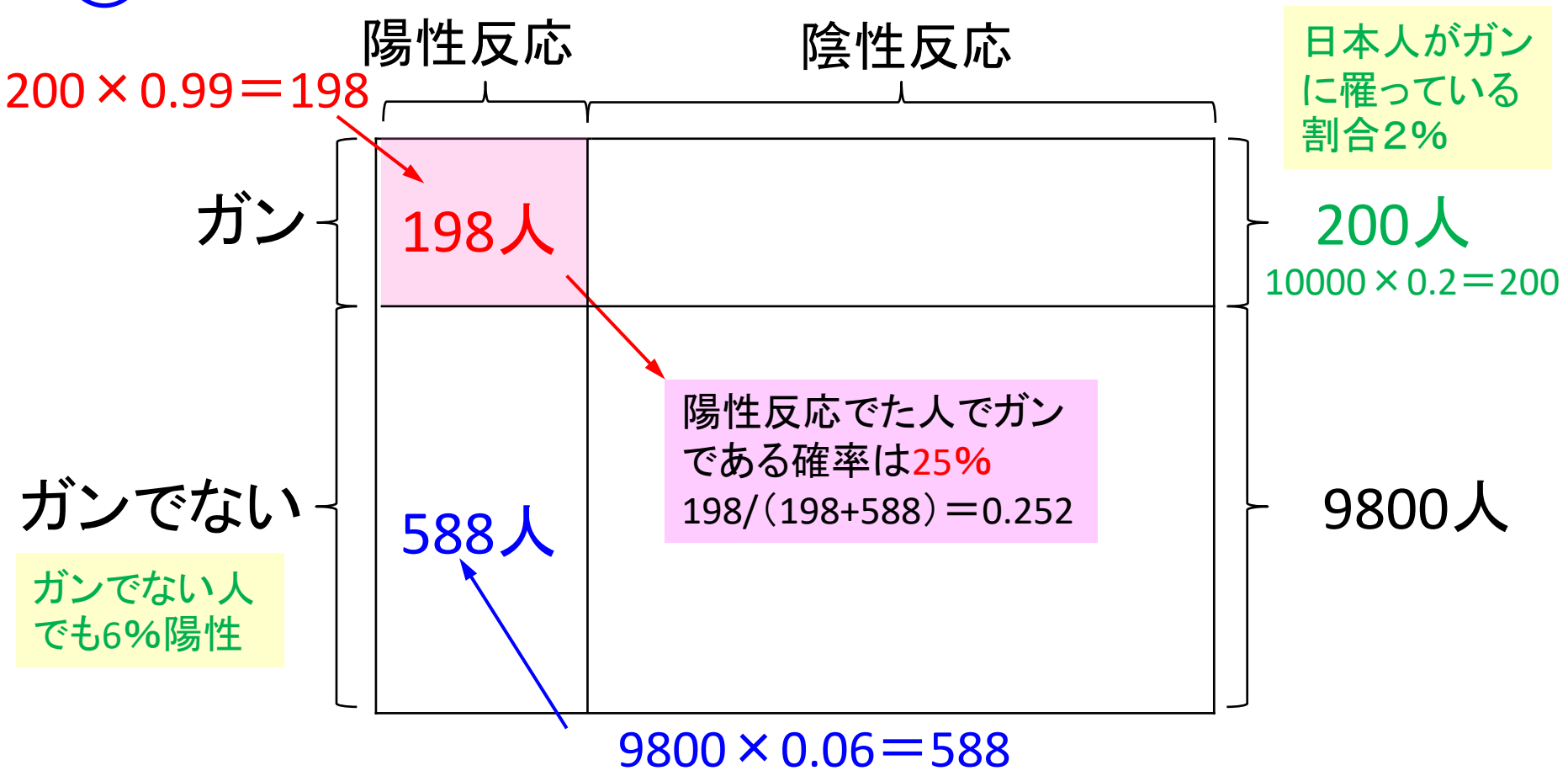
ガンに罹っている人の99%に陽性反応がでる検査がある。
太郎君がこの検査を受診したら、陽性反応がでてしまった。
太郎君がガンである確率は本当に99%なのか？

- ①検査結果から、99%の確率でガンと言える
- ②ガンであるか否かだから、50%の確率である。
- ③これだけの情報では何とも言えない。

問題12

ガンに罹っている人の99%に陽性反応がでる検査がある。
太郎君がこの検査を受診したら、陽性反応がでてしまった。
太郎君がガンである確率は本当に99%なのか？

- ①検査結果から、99%の確率でガンと言える
- ②ガンであるか否かだから、50%の確率である。
- ③これだけの情報では何とも言えない。



問題13

結婚前半の15年間と後半の10年間で夫から請求される香典の回数の変化に気づいた主婦います。以下の何れか？

- ①後半の方が香典の回数が増えた
- ②香典の度数分布の特徴が変化した
- ③最頻値(モード)が1回(または2回)から5回に変化した

結婚前半の15年間

香典回数	度数(年数)
0	1
1	4
2	4
3	3
4	2
5	1
6	0
7	0
8	0
9	0
10	0

結婚後半の10年間

香典回数	度数(年数)
0	0
1	0
2	0
3	0
4	4
5	5
6	1
7	0
8	0
9	0
10	0

問題13

結婚前半の15年間で夫から請求される香典の回数の変化に気づいた主婦います。以下の何れか？

- ①後半の方が香典の回数が増えた
- ②香典の度数分布の特徴が変化した
- ③最頻値(モード)が1回(または2回)から5回に変化した

結婚前半の15年間

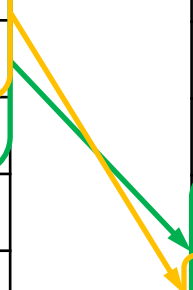
香典回数	度数(年数)
0	1
1	4
2	4
3	3
4	2
5	1
6	0
7	0
8	0
9	0
10	0

年平均
2.26

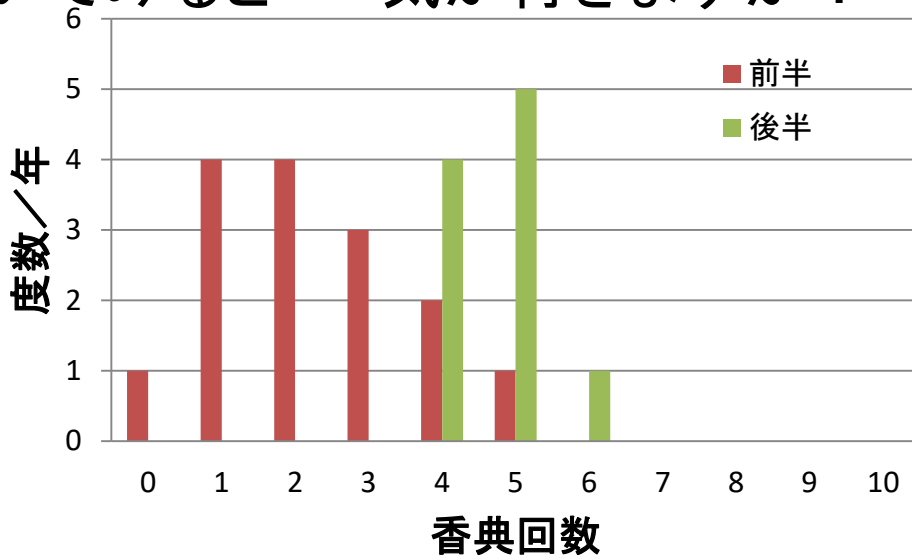
結婚後半の10年間

香典回数	度数(年数)
0	0
1	0
2	0
3	0
4	4
5	5
6	1
7	0
8	0
9	0
10	0

年平均
4.7

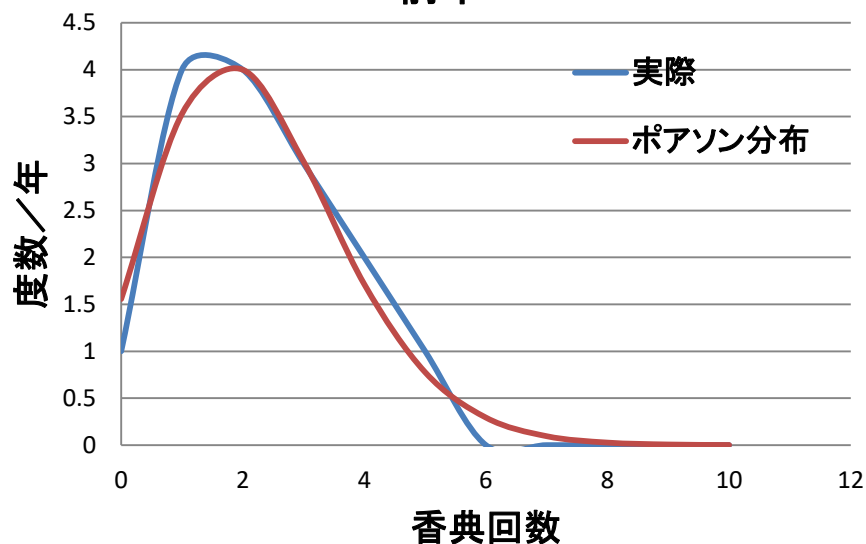


度数分布表を描いてみると・・・ 気が付きますか？

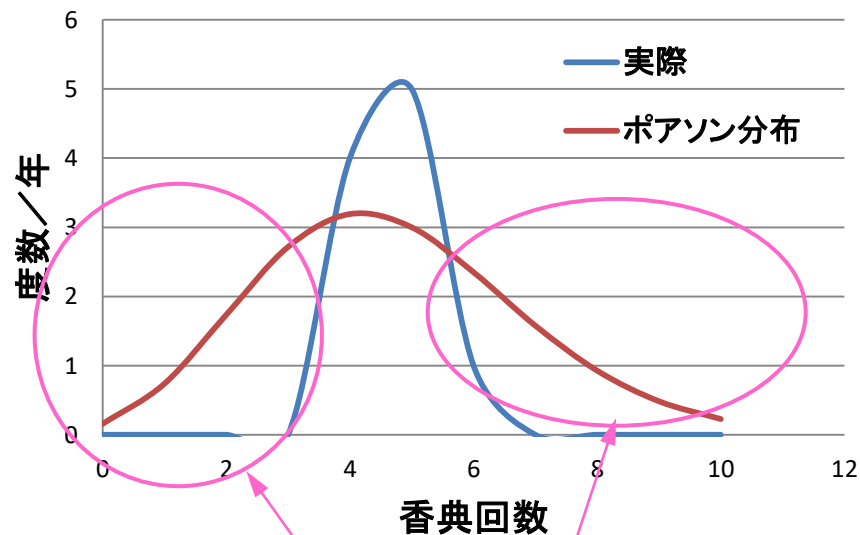


稀にしか起こらないものは、ポアソン分布に従います。

前半



後半



この辺の度数が少ない

問題14

ある都市の降水日数は年間112日。曜日ごとの降水日数は下表のとおり。土曜日は水曜日より3割以上も雨が多い。

- ①3割以上も多い曜日があるので、偏っている。
- ②この程度の差では、偏っているとは言えない。

	日曜	月曜	火曜	水曜	木曜	金曜	土曜	合計
降水日数	15	16	17	14	15	16	19	112

問題14

ある都市の降水日数は年間112日。曜日ごとの降水日数は下表のとおり。土曜日は水曜日より3割以上も雨が多い。

- ①3割以上も多い曜日があるので、偏っている。
- ②この程度の差では、偏っているとは言えない。

	日曜	月曜	火曜	水曜	木曜	金曜	土曜	合計
降水日数	15	16	17	14	15	16	19	112

仮定H 雨は曜日に関係なく均等に降る。

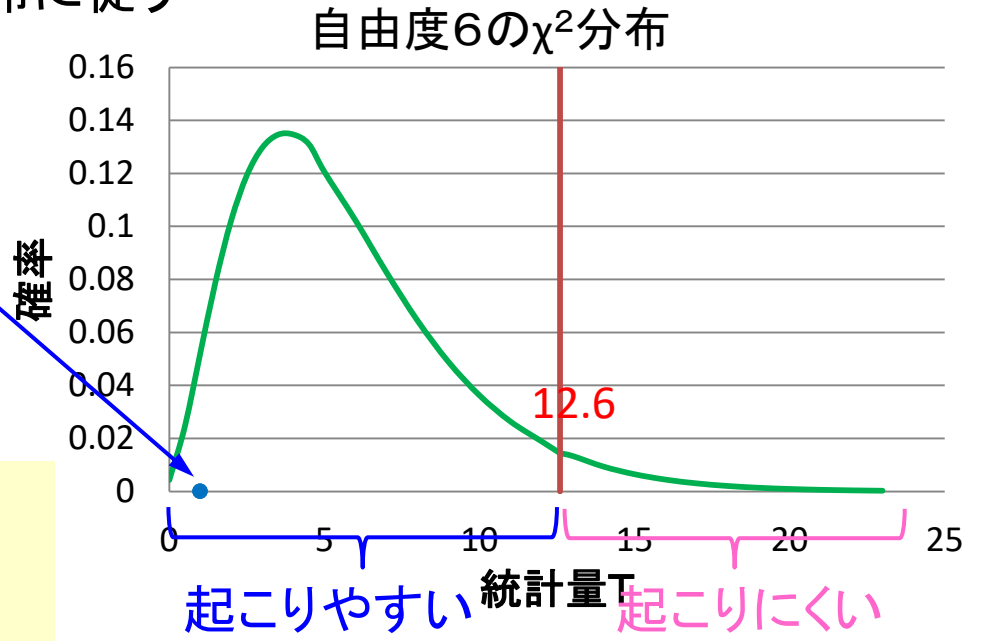
降水日数	16	16	16	16	16	16	16	112
------	----	----	----	----	----	----	----	-----

統計量Tは自由度6(=7-1)のχ²分布に従う

T = 各曜日の $\frac{(\text{観測度数} - \text{理論度数})^2}{\text{理論度数}}$ の総和

$$T = \frac{(15-16)^2 + (16-16)^2 + \dots + (19-16)^2}{16} = 1$$

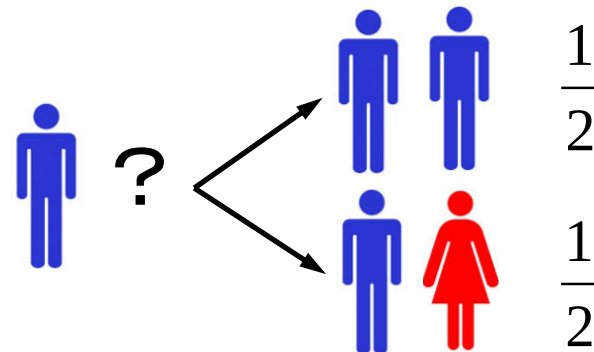
曜日に関係がなければTは小さくなる。
Tが12.6以上になる確率は5%以下
T=1なので、上記仮定Hは受容可



問題15

ある家族には、子供が2人。男の子が1人いることが判明した。
残りの1人が女の子である確率はどのくらいか？

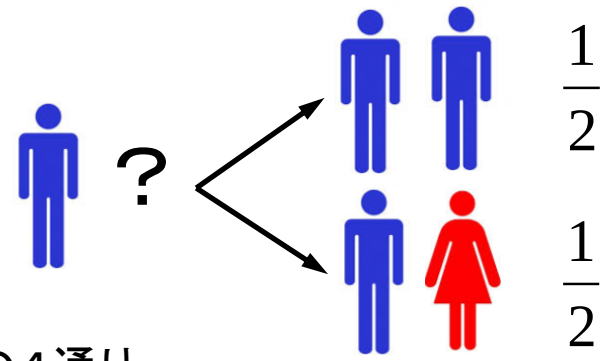
- ① $1/2$
- ② $1/2$ より大きい
- ③ $1/2$ より小さい



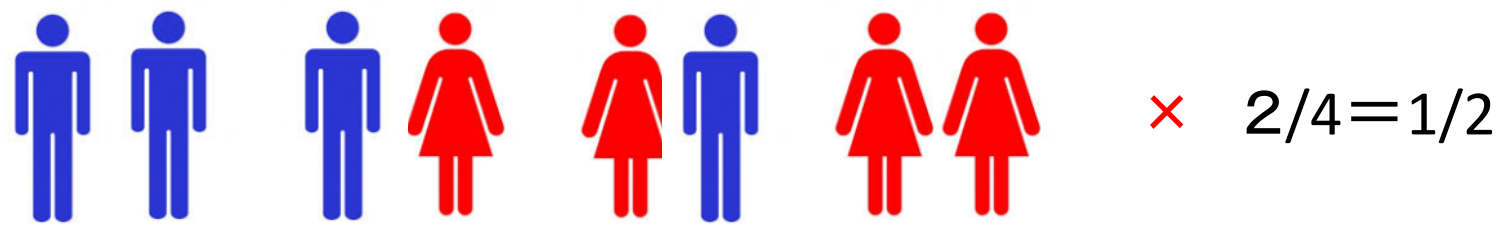
問題15

ある家族には、子供が2人。男の子が1人いることが判明した。
残りの1人が女の子である確率はどのくらいか？

- ① 1/2
- ② 1/2より大きい
- ③ 1/2より小さい



男子と女子が公平に生まれてくるとすると、以下の4通り



男の子が1人いることが判明した時点で



残りの1人が女の子である確率は、
3通り中2通りなので
2/3である

問題16

開票率1%で、なぜ「当確」を出せるか？ その理由は？

投票者が10万人いて、1000人分(1%)を開票した時、立候補者Aの得票率が0.6であるとして説明せよ。

問題16

開票率1%で、なぜ「当確」を出せるか？ その理由は？

投票者が10万人いて、1000人分(1%)を開票した時、立候補者Aの得票率が0.6であるとして説明せよ。

開票数n、その中で立候補者Aの得票率をrとする。真の得票率Rは95%の信頼度では以下の式となる。

$$r - 1.96 \times \sqrt{\frac{r(1-r)}{n}} \leq R \leq r + 1.96 \times \sqrt{\frac{r(1-r)}{n}}$$
$$0.6 - 1.96 \times \sqrt{\frac{0.6(1-0.6)}{1000}} \leq R \leq 0.6 + 1.96 \times \sqrt{\frac{0.6(1-0.6)}{1000}}$$

$0.57 \leq R \leq 0.63$

開票率	開票数	推定得票率下限	推定得票率上限
1%	1,000	0.570	0.630
2%	2,000	0.579	0.621
5%	5,000	0.586	0.614
10%	10,000	0.590	0.610
15%	15,000	0.592	0.608
20%	20,000	0.593	0.607
25%	25,000	0.594	0.606
30%	30,000	0.594	0.606
35%	35,000	0.595	0.605
40%	40,000	0.595	0.605
45%	45,000	0.595	0.605
50%	50,000	0.596	0.604
55%	55,000	0.596	0.604
60%	60,000	0.596	0.604
65%	65,000	0.596	0.604
70%	70,000	0.596	0.604
75%	75,000	0.596	0.604
80%	80,000	0.597	0.603
85%	85,000	0.597	0.603
90%	90,000	0.597	0.603
95%	95,000	0.597	0.603
100%	100,000	0.597	0.603

