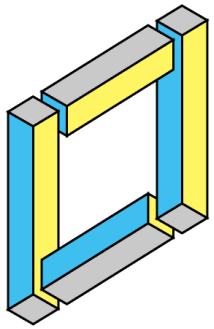
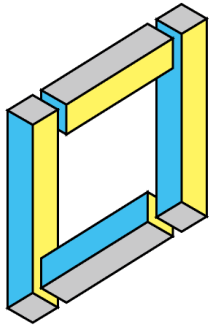
データ分析以前の着眼点

そもそも統計とは何か ① 統計の基本的な姿勢



統計とはなにか？



はじめに（１）

以下の数値において、統計としての数値に当てはまるか考えてみましょう。また、そのように判断した理由を考えましょう。

- **中学生10人の平均身長**
- **2022年、生産人口世代の年収の中央値**
- **世界大学ランキングの順位**
- **各政党の支持率**

はじめに（２）

・中学生10人の平均

・2022年

・

・



はじめに（３）

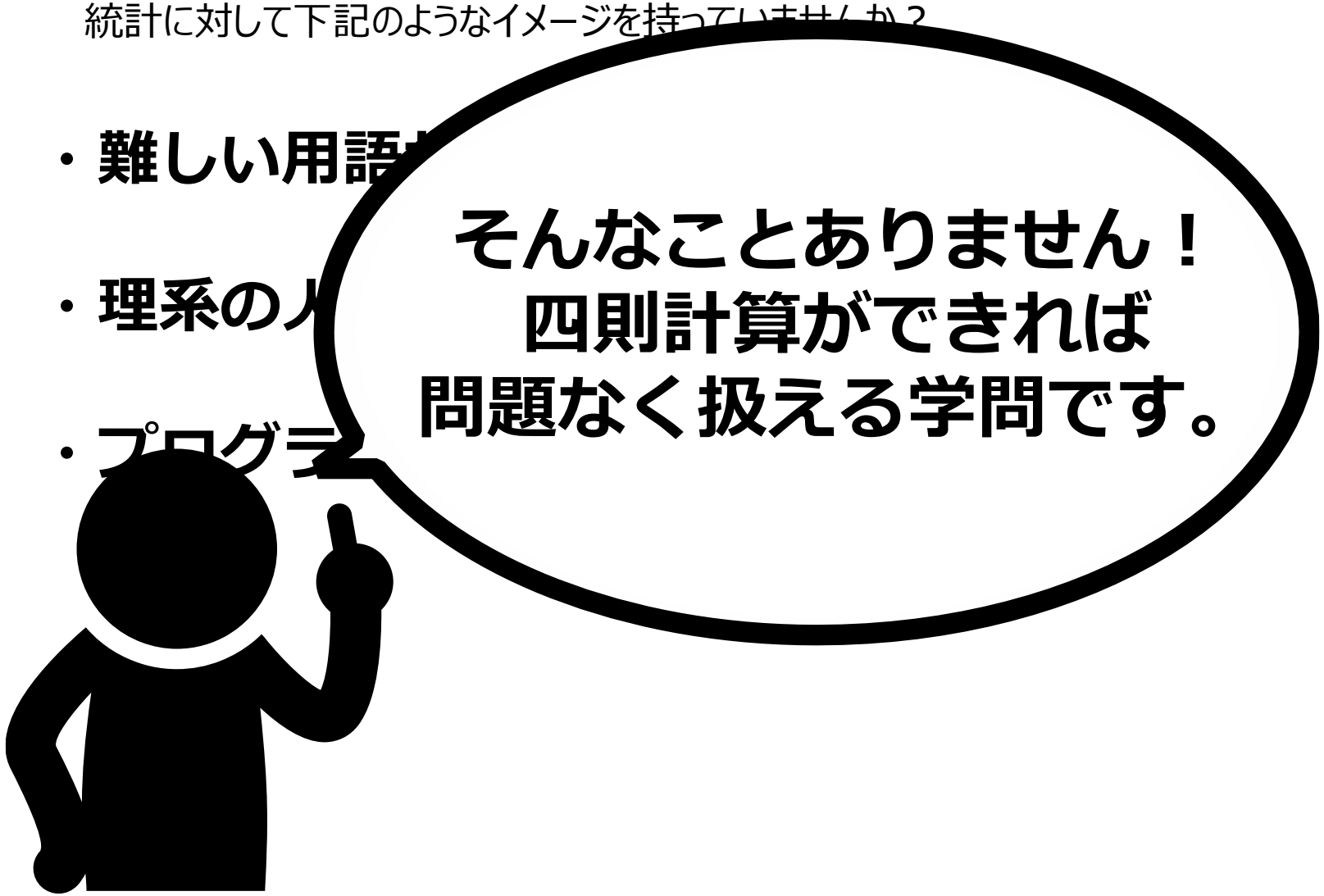
統計に対して下記のようなイメージを持っていませんか？

- **難しい用語が並んでいる学問**
- **理系の人しか扱えない**
- **プログラミングや数学の知識が必要**

はじめに（４）

統計に対して下記のようなイメージを持っていませんか？

- **難しい用語**
- **理系の人**
- **プログラ**



統計とは何か？（１）

統

まとめる

多くのものを
１つにまとめる

計

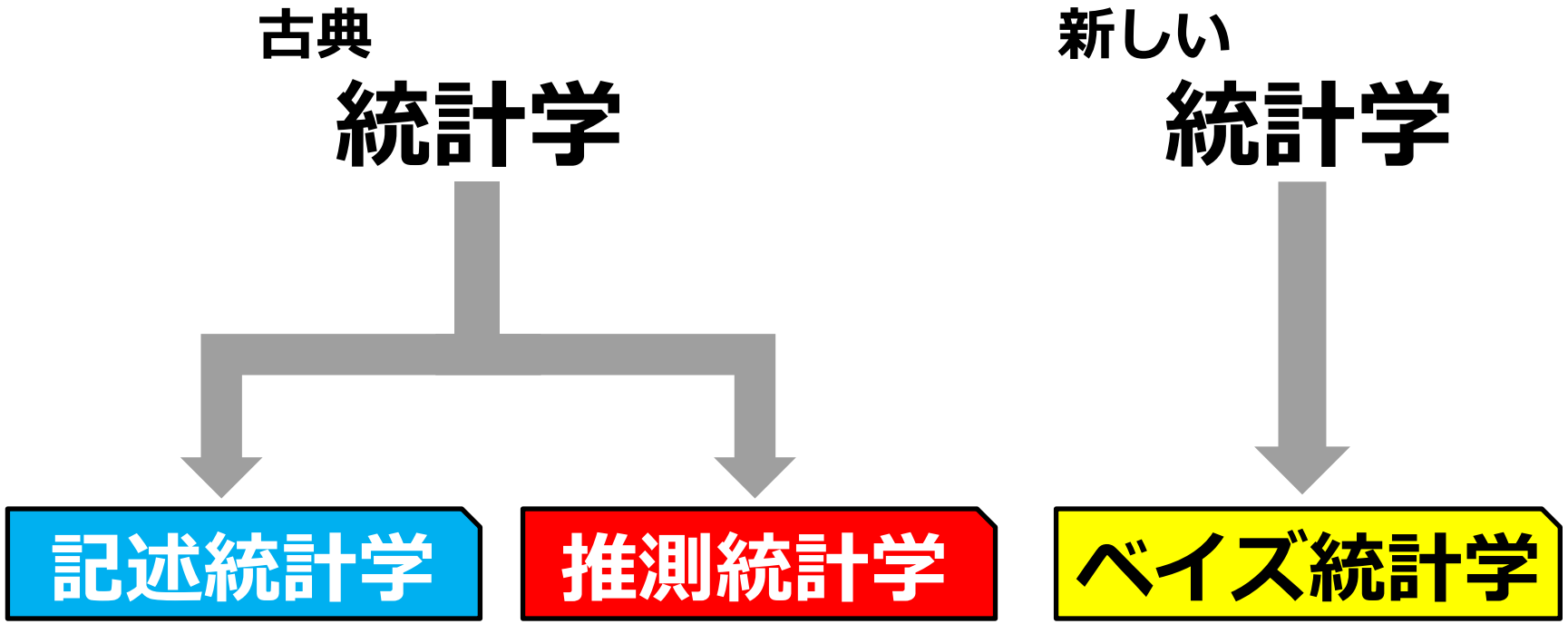
はかる

ものの数値を調べる
(重さや長さなど)

全てをあつめて計算する

統計とは何か？（２）

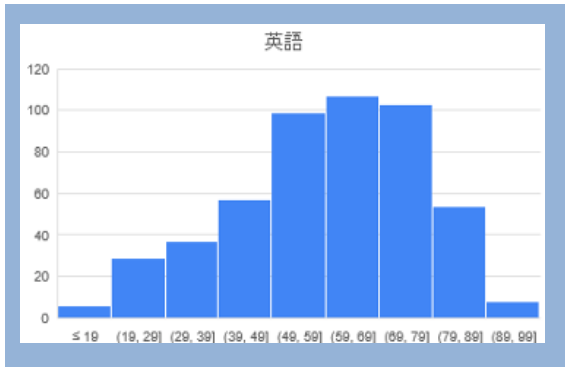
統計には大きく分けて３種類あります。



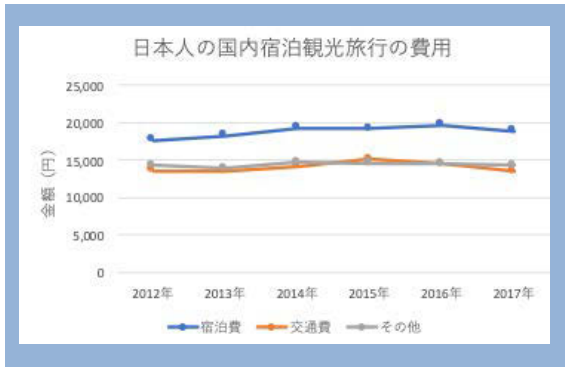
記述統計学とは（１）

データを集約して、表やグラフなど見やすい形に整え
データの特徴を捉えること

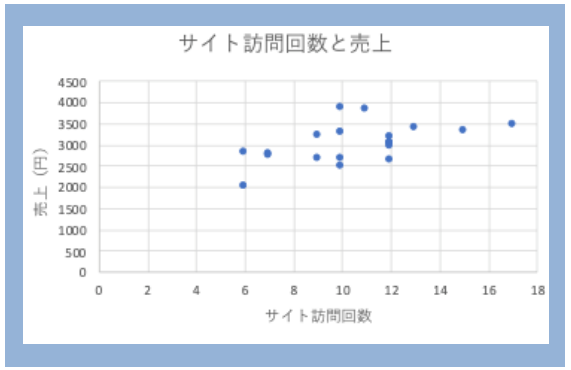
→人に何かデータを伝えるときに使用する



ヒストグラム



折れ線グラフ



散布図

記述統計学とは（２）

例：どちらの数値の方が大きいと感じますか？

A　{2,5,10,25,50,100}

B　{5,30,40,50,60,70}

Aには最大値の100があるため、大きい？

Aには最小値２があるため、小さい？

Bは各要素の合計値が大きいため大きい？

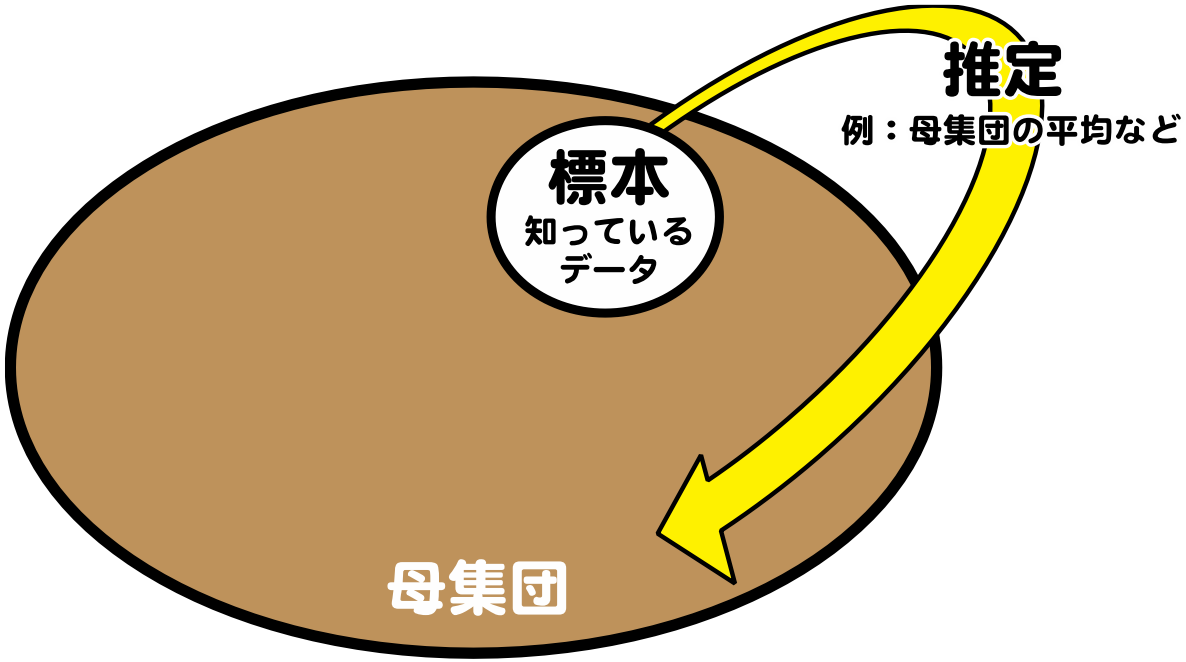
Bには３桁数値がないため、小さい？

今回は数値が少ないのでイメージしやすいが、
これが１万件のデータがあった場合、どちらが大きいか判断できますか？
それを表やグラフで可視化するのが記述統計学です。

推測統計学とは（１）

必要なデータを全部持ってなかったとしても、
持っているデータから全体の傾向を推測すること

→ビジネスシーンでよく使用される



推測統計学とは（２）

例：あるPC企業は

・80cmの落下耐久があります
・キーボードが10万回押しても壊れません
・70度の気温下で使用しても壊れません

としてPCを販売していますが、
販売するPCの全てをチェックすることは可能でしょうか？

→　無作為に５千台程度に１台抽出してチェックを行い、
３つの試験に合格しているため、このように記載が可能になります。

ベイズ統計とは

今あるデータのみから母集団を考える統計学

→推測統計よりも主観的にデータを扱いながら統計を進められる



通常の統計学

母集団／真の確率は
決まっている

例：コインを投げる

実際に投げて観察する

表が何回出たという事象

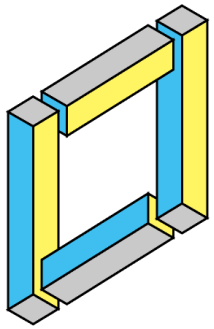
ベイズ統計

特定の事象がある

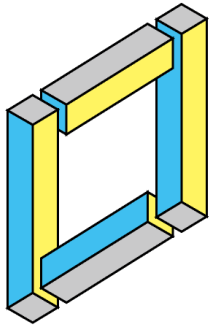
例：コインを投げた時に
表が何回出たという事象

主観的に推測する

コインを投げると表が
この確率で出るだろう

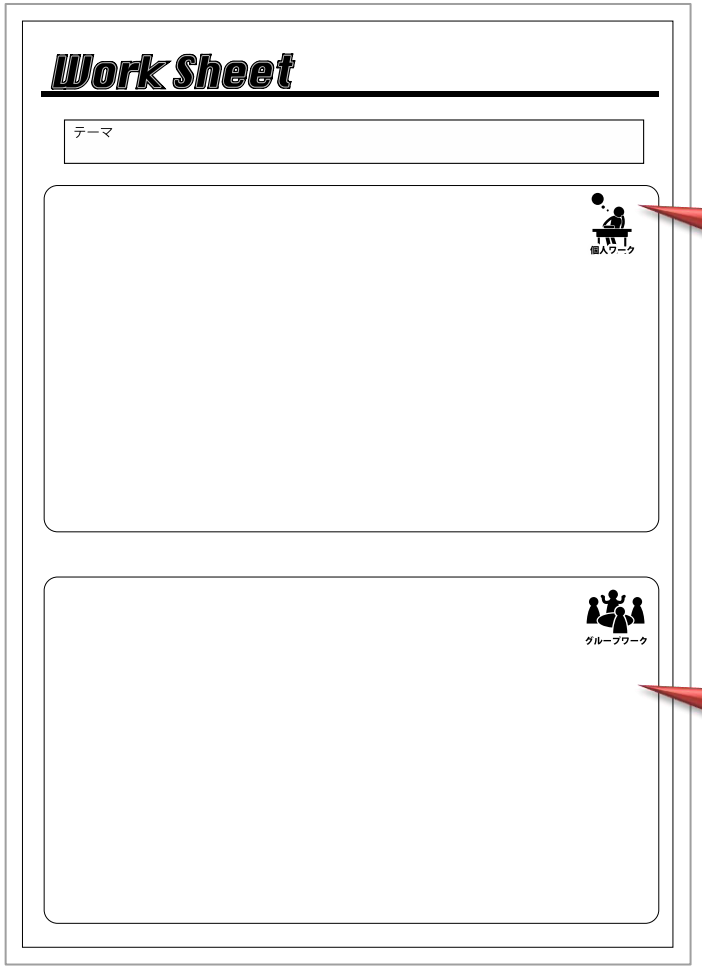


ワーク



統計ではないもの

身の回りの数値や、データで統計ではないものにはどんなものがあるか考えてみましょう。
また、それはどうして統計ではないと考えた理由をグループで共有してみましょう。

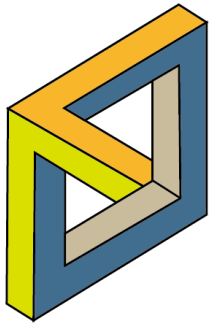


これは違うかも？
と思ったものでも、どんどん書き出していきましょう。
書き出した後に、統計かそうでないかを考えましょう。

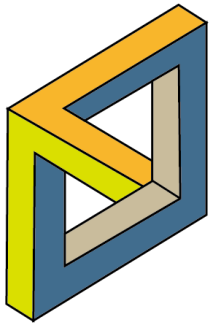
考えたものをグループで共有してましょう。
なぜ、統計ではないと思ったのかをグループメンバーに説明しましょう。

データ分析以前の着眼点

そもそも統計とは何か ② 統計でわかること・わからないこと



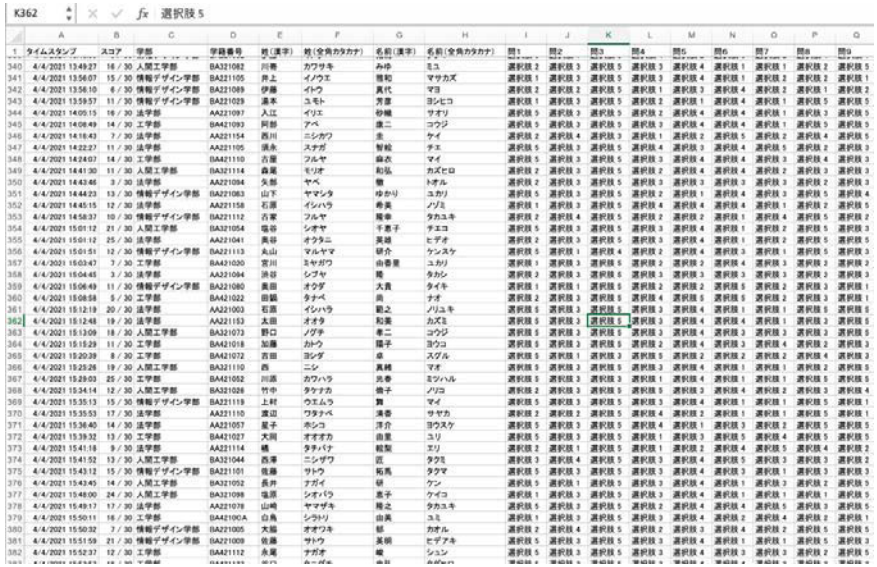
統計でわかること・わからないこと



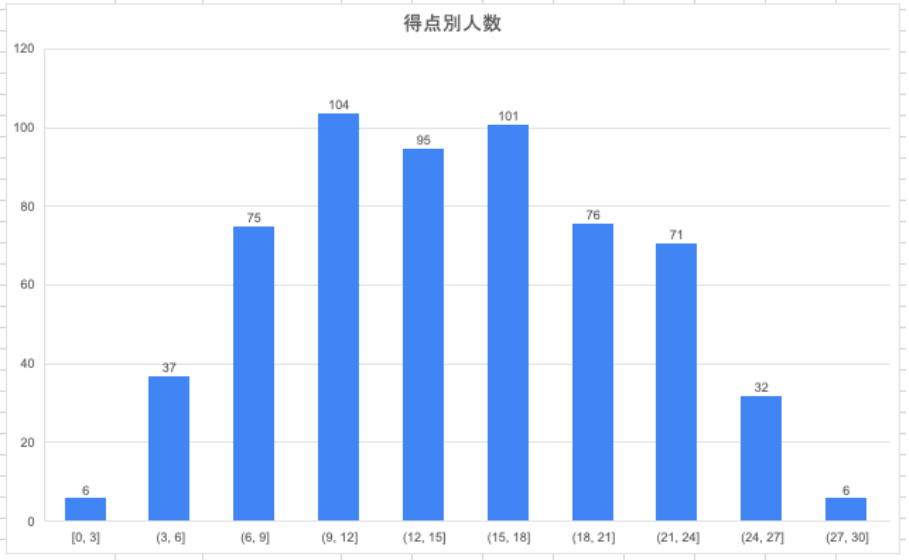
統計が得意とすること

大量かつ複雑なデータを人間が見て理解しやすくする

Excelの一覧表

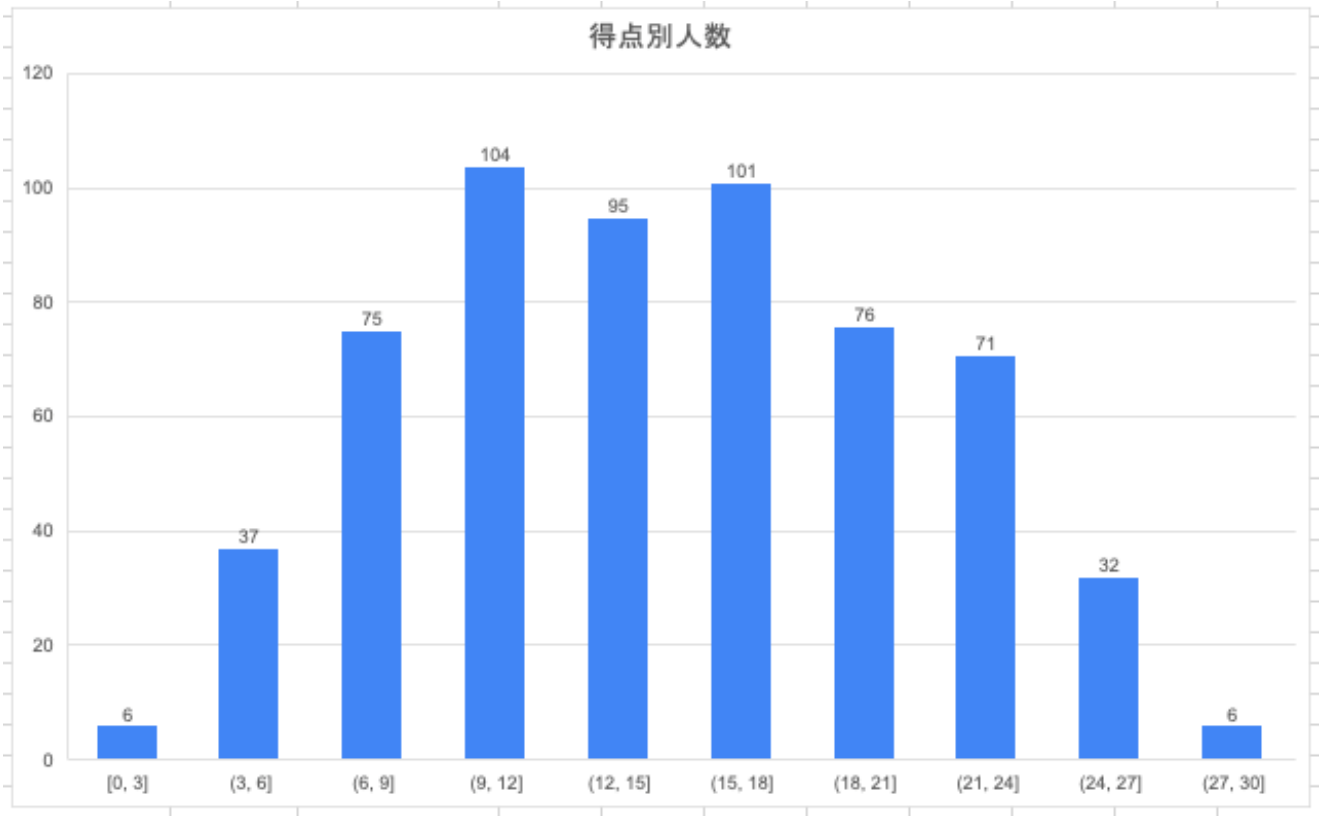


分析したいデータをヒストグラム化



統計ではできないこと

データ(結果)を見て考察するのは自分



このグラフからわかることはなんだろうか？

統計を行う上での注意点（１）

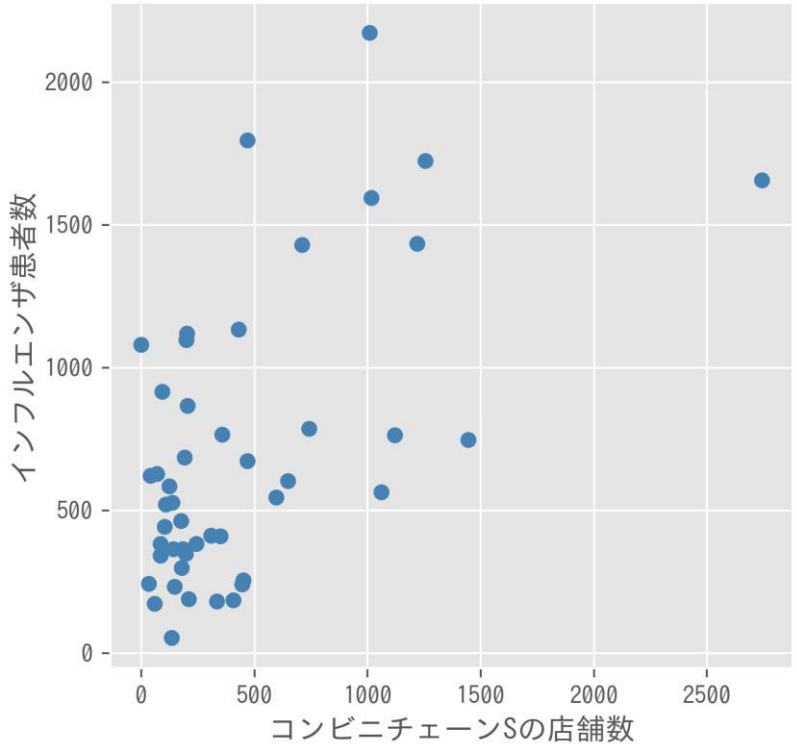
見た目に騙されないようにする



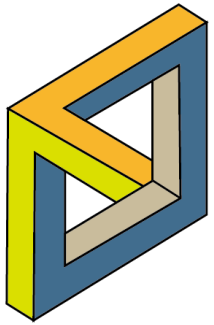


統計を行う上での注意点（２）

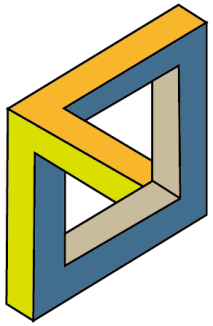
結果から間違った方向性に向かないようにする



左の図はインフルエンザ患者数と
コンビニチェーン店の店舗数の散布図です。
コンビニチェーン店が多いとインフルエンザ
の患者数が多くなるようなので、
コンビニの数を減らせばインフルエンザの
患者数は減る…？

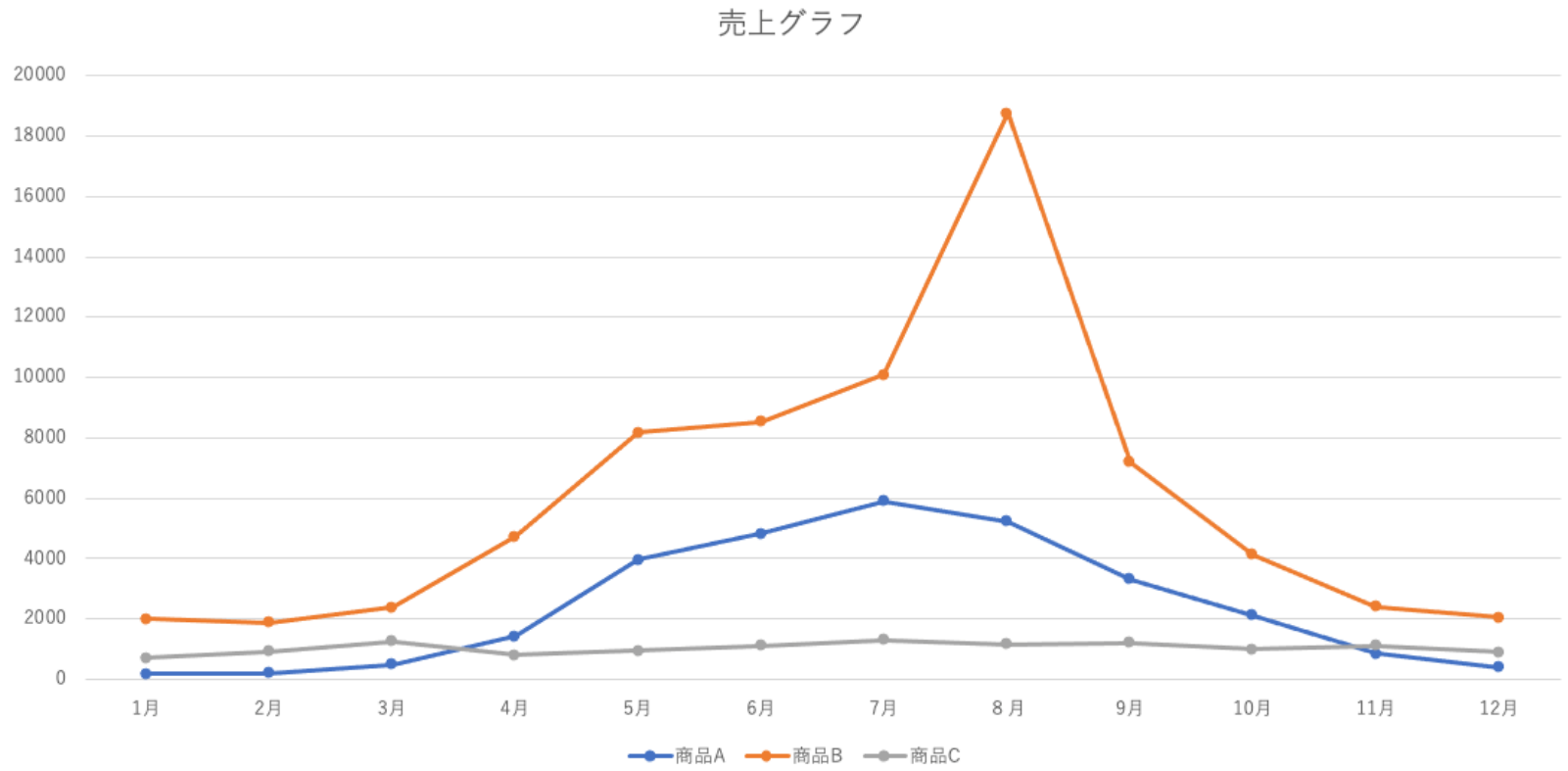


ワーク

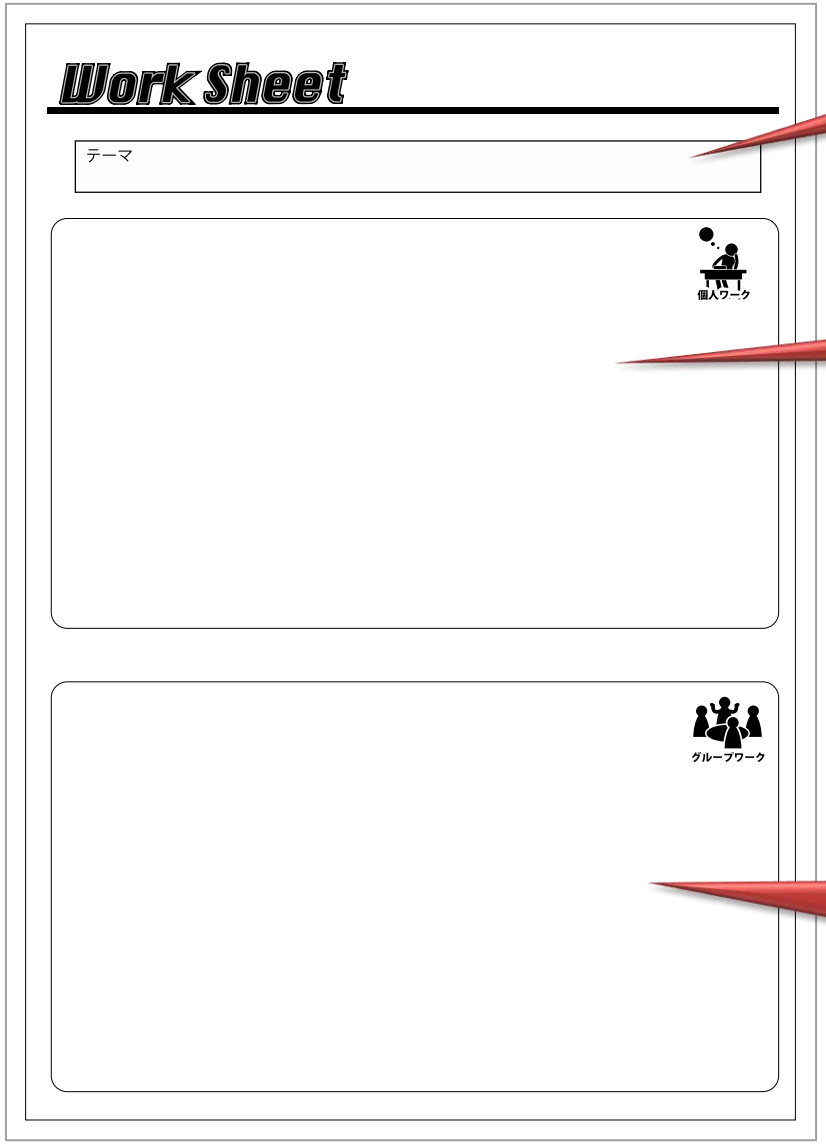


ワーク（1）

以下のグラフからわかることを書き取りなさい。



ワーク（２）



テーマを書き込む

グラフを見て、直感的に分かることから
どんどん書き出していきましょう。

何を読み取れたかをグループで話し合
いましょう。
グループメンバーの意見で、いいと
思ったことは、どんどん取り入れてい
きましょう。

ワーク（3）

グループごとの意見を全体に共有しましょう。

どんなことを読み取れましたか？

※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。



データ分析以前の着眼点

事実と意見①　事実と意見の違い・代表的な混同

事実と意見の違い

事実：**真偽が証明できる**
⇨誰もが同意する

人間には、水が必要だ

意見：**真偽が証明できない**
⇨その人の主観による

水を飲むと元気が湧く

事実と意見の区別

円安が大幅に加速し、１ドル140円台にまで進行しました。

１ドル140円台に→ **真偽が確認できるので「事実」**

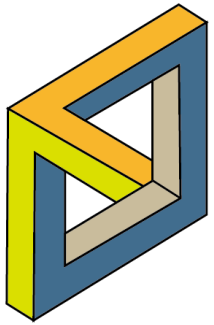
大幅に加速し → **「大幅」か否かは個人の主観なので「意見」**

事実と意見の区別

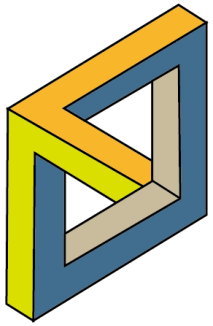
4月30日の国内新型コロナウィルスの感染者は、
2万5182人で以前としてパンデミックは継続傾向にある。

感染者は2万5182人→ **真偽が確認できるので「事実」**

継続傾向 → **データを元にしてはいるが、解釈なので「意見」**



ワーク



話し合いのなかの事実と意見

これから、グループでディスカッションをしてもらい、その中での意見と事実の使われ方を振り返ります。

クラスを大きく2つ（Aサイド・Bサイド）に分けます。

→そのなかでさらに4-5人のグループになってもらいます。

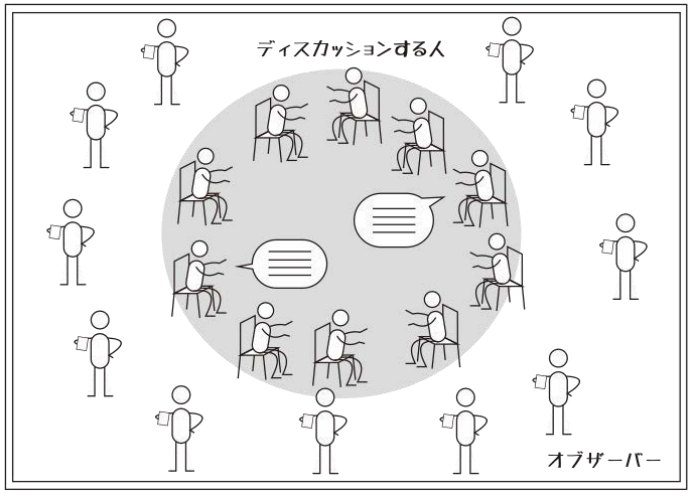
Aサイドのグループがディスカッションをしている間、Bサイドのグループは観察します。

Bサイドのグループがディスカッションをしている間、Aサイドのグループは観察します。

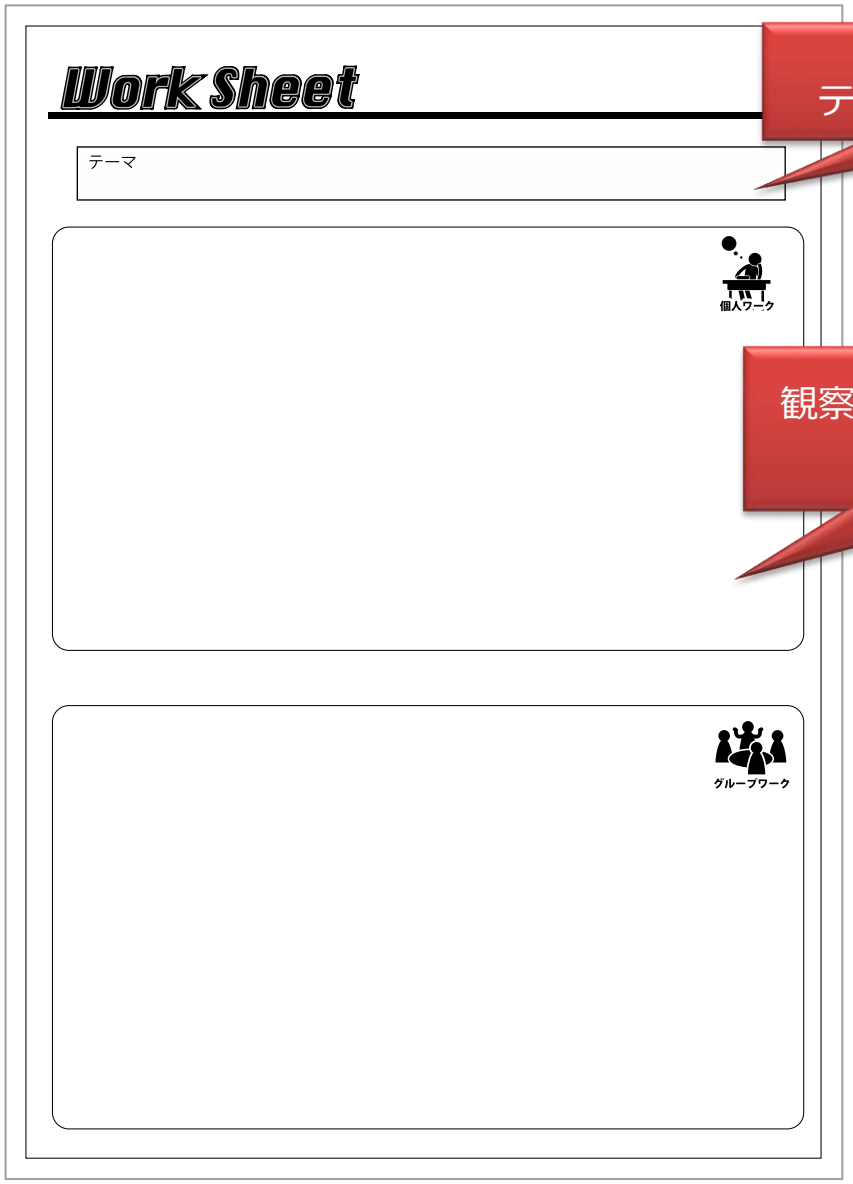
観察の中では、下記を行ってください。

①発言内容を記録する（特に主張の根拠）

②その発言が事実なのか意見なのかを分類する。



話し合いのなかの事実と意見



後で見て分かるように
テーマを書いておきましょう

観察対象グループの発言を記録
※書き取れる範囲で

記録しながら、発言を分類します。

事実＝客観的で真偽判定可能
意見＝主観的で審議判定が不可能

分類例
　事実を四角、意見を丸で囲む
　事実に☆マーク、意見に○をつける

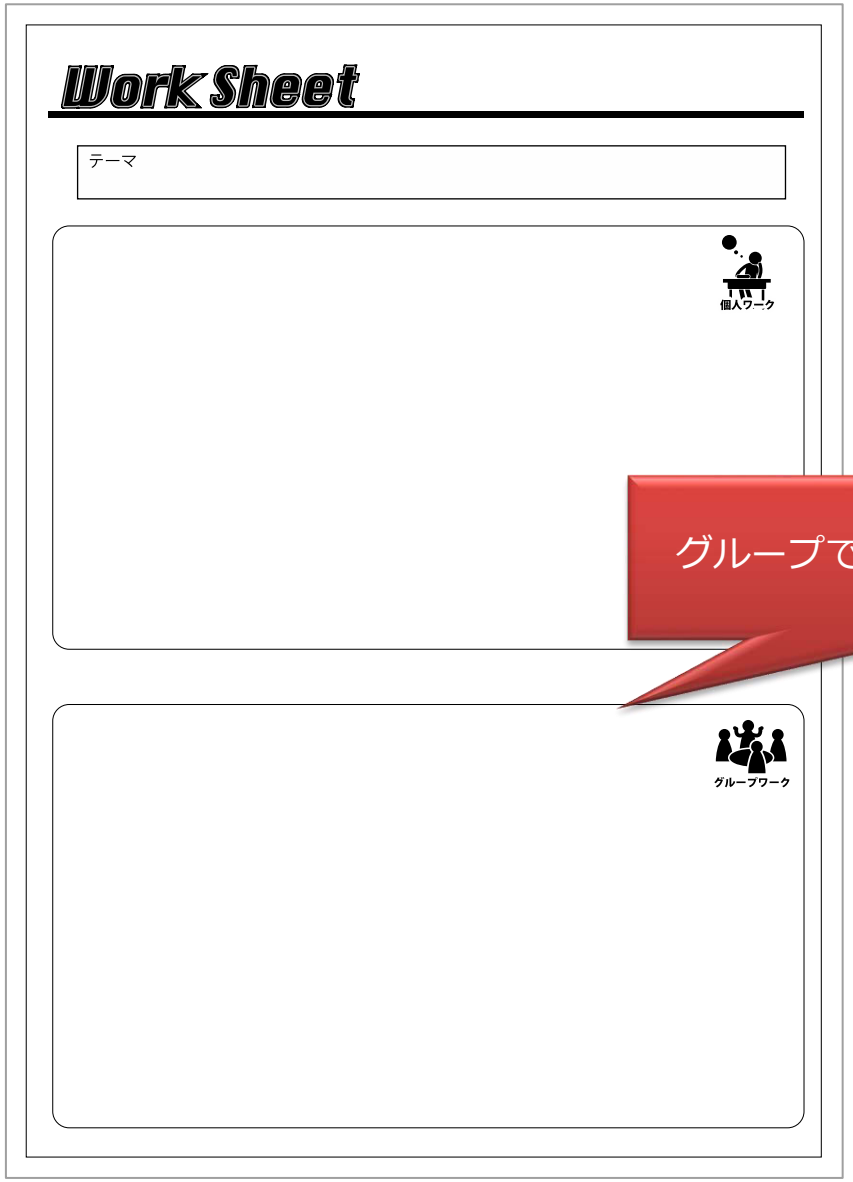
Aサイド用テーマ１

ライフラインの整備されていない無人島で１年間
サバイバル生活をしなくてはならなくなりました。
島には、命を脅かす猛獣や毒を持つ危険な生物は
いません。きれいな水が流れている川もあります。
家から何かアイテムを３つ持って行っていいこと
になっているときに、何を持っていきますか？

Bサイド用テーマ１

もはや日本を代表するメニューのカレーライスと
ラーメン。
アフリカのテレビ局で、カレーライスかラーメンを
紹介するテレビ番組の企画を担当することになった
チームのメンバーとして、カレーライスかラーメン
のどちらを選ぶか討論し、決定してください。
ただし、アフリカのどこの国で放送されるかは、
まだ知らされていません。

話し合いのなかの事実と意見



グループで意見交換をしましょう。

個人での記録と分類をグループ内で共有し
グループとしての観察結果を考えてください。
なるべく結論がでることを目指してください。

話し合いのなかの事実と意見

グループごとの観察結果を全体に共有しましょう。

共有の観点

①どんな発言があったか

②事実と意見の比率は、どれくらいか

③どちらの方が根拠として効果的に作用していたと思うか

※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。



なぜ、事実と意見を混同するのか？

〇原因①：話し手自身が事実と意見を混同している

「富士山は、世界で一番美しい山だ」

「今の政治はダメだ」

⇨発言者は事実だと思っており、悪意はない

〇原因②：あえて事実と意見を混同するように書いている

「カップメンを食べた人の死亡率は100%です」

→100年後・200年後であれば、死亡して当然である

「弊社の8割の製品に最新の環境基準が適応されています。」

→実は残りの2割が主力消費で全体の生産高の8割を担っている

⇨悪意有り（印象操作により受け手が誤解するように仕向けている）

事実と意見の区別

自分の発言や報告で、無意識に事実と意見を混同しないようにしましょう。

事実と意見を意識的に混同させようとしている発言や文章に気を付けましょう。

データ分析以前の着眼点

事実と意見②　事実と意見を分ける演習

事実と意見を分けて考える

以前に分析した下記結果を元に

事実と意見を分けて考える練習をしましょう。

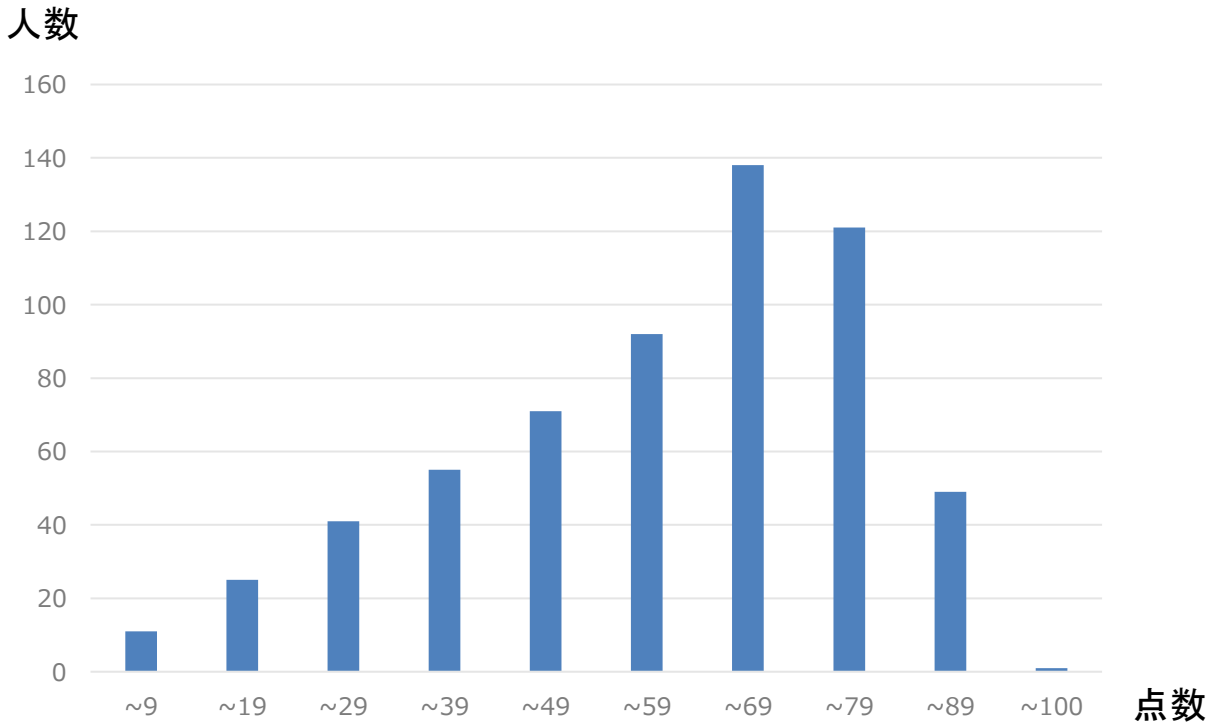
	全体	A学科	B学科	C学科	D学科
平均値	56	75	65	51	29
中央値	60	76	65	50	29
最頻値	71	67	71	50	42
標準偏差	20	7	8	11	13
平均値と中央値の差（絶対値）	4.4	1.1	0.2	0.8	0.0

①結果から読み取れる事実をあげていってください。

②そこから導き出せる意見（解釈）をあげてください。

事実と意見を分けて考える

下記結果を元に事実と意見を分けて考える練習をしましょう。



①グラフから読み取れる事実をあげていってください。

②そこから導き出せる意見（解釈）をあげてください。

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

分析結果から読み取れることを
整理していきましょう。

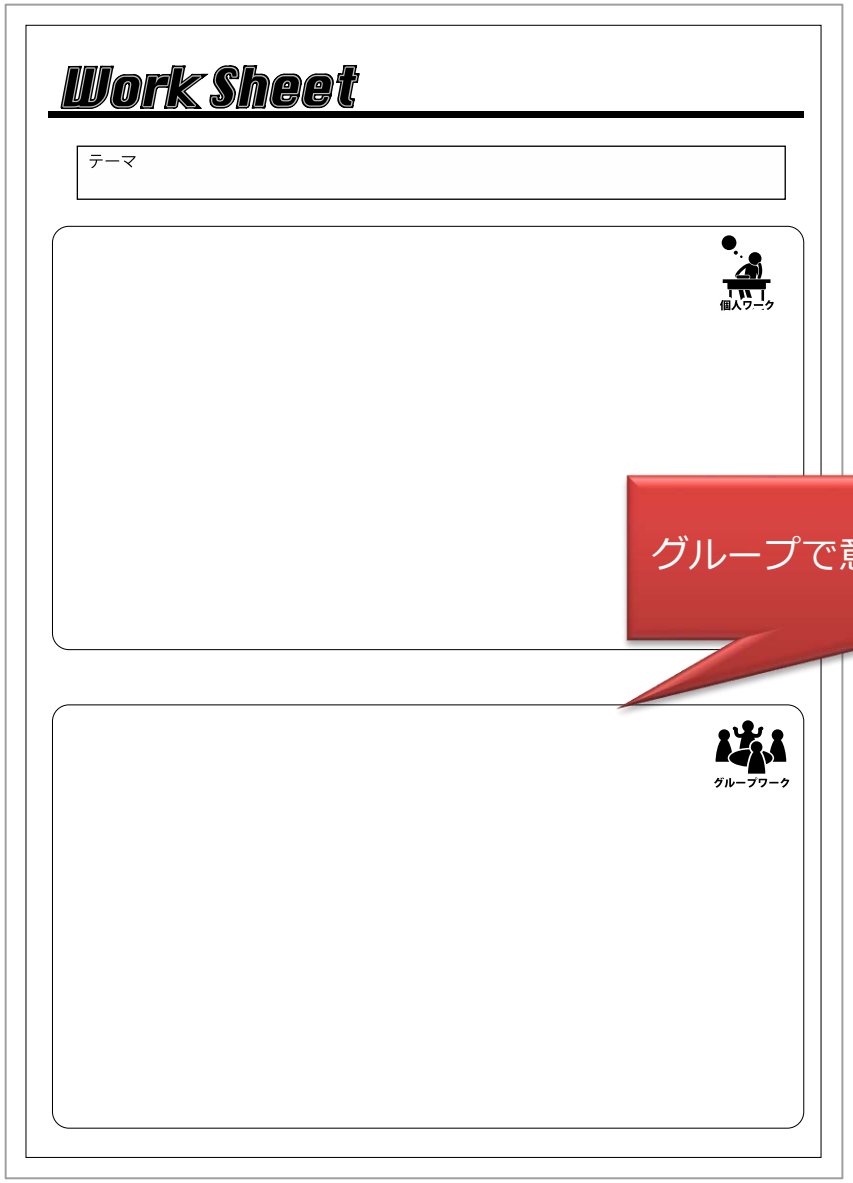
記入しながら事実と意見（解釈）を分類します。

分類例

事実を四角、意見を丸で囲む

事実に☆マーク、意見に○をつける

ワークの進め方（2）



グループで意見交換をしましょう。

個人での読み取りと分類をグループ内で共有し
グループとしての解釈を考えてください。
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・解釈」はないか
議論して探ってみましょう。

ワークの進め方（3）

グループごとの観察結果を全体に共有しましょう。

共有の観点

①どんな事実が見つかったか

②どのような解釈が考えられるか

③その根拠はなにか

※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。



解答例

	全体	A学科	B学科	C学科	D学科
平均値	56	75	65	51	29
中央値	60	76	65	50	29
最頻値	71	67	71	50	42
標準偏差	20	7	8	11	13
平均値と中央値の差（絶対値）	4.4	1.1	0.2	0.8	0.0

①結果から読み取れる事実をあげていってください。

1）平均点は、A学科＞B学科＞C学科＞D学科となっているが、　標準偏差は逆の順序になる

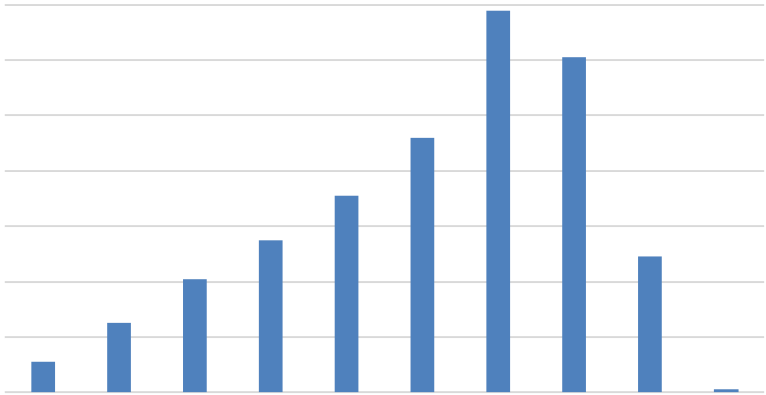
2）平均値と中央値の差はA学科が一番大きく、D学科が一番小さい

②そこから導き出せる意見（解釈）をあげてください。

1）より、平均点の高い学科ほど、学生の差が大きいと考えらえる（標準偏差が大きい）

2）より、平均点の高い学科は一部の学生が全体の平均を押し上げている可能性がある

解答例



①結果から読み取れる事実をあげていってください。

1）60～69点が最も多い

2）90~100点より、0~9点の方が多くいる

②そこから導き出せる意見（解釈）をあげてください。

1)2)より、全体的に高得点（50点より上）に偏っている

データ分析以前の着眼点

確率（確率の計算）① 確率計算の基本

乗法定理と加法定理

乗法定理と加法定理

確率の計算をする上で、乗法定理と加法定理の意味をきちんと理解しておくことは重要です。乗法定理と加法定理は一般的に、積の法則、和の法則とも呼ばれることがあります。

乗法定理

お互いに影響を与えない、独立した事象が連続した時の確率を求める場合は、それぞれの事象の確率の積を求める。

加法定理

ある事象で、どちらか一方が起きるともう片方が起きえない（同時には起こりえない事象）で、どちらかが起きる確率を求める場合は、それぞれの事象の和を求める。

言葉の説明を聞いても、イメージしにくいので、具体例を交えて見ていきましょう。

乗法定理

サイコロを２つふった時、２つとも１の目が出る確率は？

２つのサイコロをサイコロＡとサイコロＢとしましょう。この２つのサイコロの出る目は、お互いに影響しあいません。ですから、サイコロＡをふった時に１が出る確率とサイコロＢをふった時に１が出る確率は、それぞれ独立した事象の連続と考えることができるので、それぞれの確率の積で求めることができます。

サイコロＡで１が出る確率

$$1 \div 6 = \frac{1}{6}$$

×

サイコロＢで１が出る確率

$$1 \div 6 = \frac{1}{6}$$

かけ算で求める！

$$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

加法定理

サイコロを２つふった時、どちらか一方が１になる確率は？

どちらか一方が１になるということは、サイコロＡをふった時に１が出た場合には、サイコロＢでは１でない目が出ています。逆に、サイコロＢで１が出た場合には、サイコロＡでは１でない目が出ています。つまり、どちらか一方が１の場合に、もう片方が１でないと事象の計算をする場合は、たし算を使います。

サイコロＡで１が出て、
サイコロＢで１以外が出る確率

サイコロＡ　　サイコロＢ

$1 \div 6 = \frac{1}{6}$　　$5 \div 6 = \frac{5}{6}$

$\frac{1}{6} \times \frac{5}{6} = \frac{5}{36}$

＋

サイコロＡで１以外が出て、
サイコロＢで１が出る確率

サイコロＡ　　サイコロＢ

$5 \div 6 = \frac{5}{6}$　　$1 \div 6 = \frac{1}{6}$

$\frac{5}{6} \times \frac{1}{6} = \frac{5}{36}$

たし算で求める！

$\frac{5}{36} + \frac{5}{36} = \frac{10}{36} = \frac{5}{18}$

少なくとも・・・の場合

サイコロを２つふった時、少なくともどちらか一方が１になる確率は？

乗法定理と加法定理の基本を理解した上で、次の事象の確率を考えてみましょう。
サイコロＡとサイコロＢの２つのサイコロをふった時に、少なくともどちらか１方が１になる確率を求めてみましょう。

ポイント

少なくとも一方が１になるという事象を考えてみましょう。この事象は、２つのサイコロのどちらかが１になった時だけでなく、両方とも１の場合でも当てはまります。また、両方とも１になる事象と、どちらか一方が１になる事象は同時には起こりえません。ですので、求めたい確率は、両方とも１になる確率と、どちらか一方が１になる確率の和で求められます。

サイコロＡ、サイコロＢの
両方で１が出る確率

$$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

＋

サイコロＡ、サイコロＢの
どちらか一方で１が出る確率

$$\frac{5}{36} + \frac{5}{36} = \frac{10}{36} = \frac{5}{18}$$

たし算で求める！　$\frac{1}{36} + \frac{5}{18} = \frac{11}{36}$

乗法定理の簡単な例

トランプを使った確率を考えてみましょう。

ジョーカーを抜いたトランプが１セット、52枚あります。
ここから、数字の７を引く確率は $\frac{1}{13}$ です。

それでは、このトランプの山から２枚引いたときに、２枚とも７である確率はいくつか？



乗法定理の簡単な例（解）

1枚引いたときに $\frac{1}{13}$ というのは、次の計算からでてきますね。

7の枚数

トランプ全体の枚数

$$\frac{4}{52} = \frac{1}{13}$$

２枚引いたというのは、1枚引いた後に、戻さずにもう1枚ひくのと同じなので、
２枚目が７である確率は

残りの７の枚数

トランプ全体の
残りの枚数

$$\frac{3}{51} = \frac{1}{17}$$

したがって、２枚とも７である確率は

$$\frac{1}{13} \times \frac{1}{17} = \frac{1}{221}$$

加法定理の簡単な例

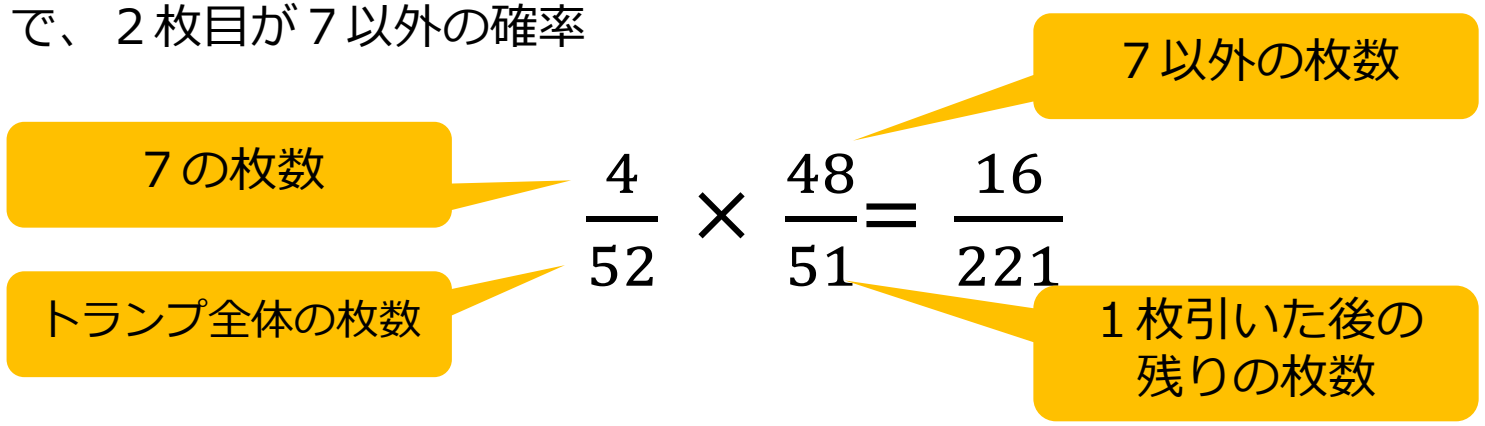
トランプを使った確率を考えてみましょう。

ジョーカーを抜いたトランプが1セット、52枚あります。ここから、数字の 7 を引く確率は $\frac{1}{13}$ です。
それでは、このトランプの山から 2 枚引いたときに、どちらか 1 枚が 7 である確率はいくつか？

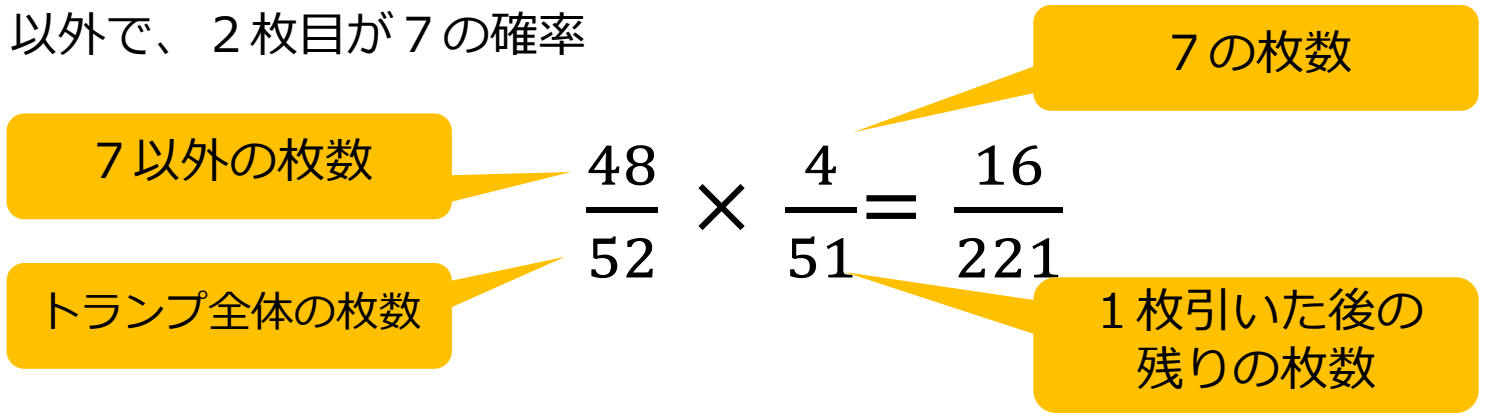


加法定理の簡単な例（解）

1枚が７で、２枚目が７以外の確率



1枚が７以外で、２枚目が７の確率



この２つの現象は同時に起こりえないので、
求める確率は

$$\frac{16}{221} + \frac{16}{221} = \frac{32}{221}$$

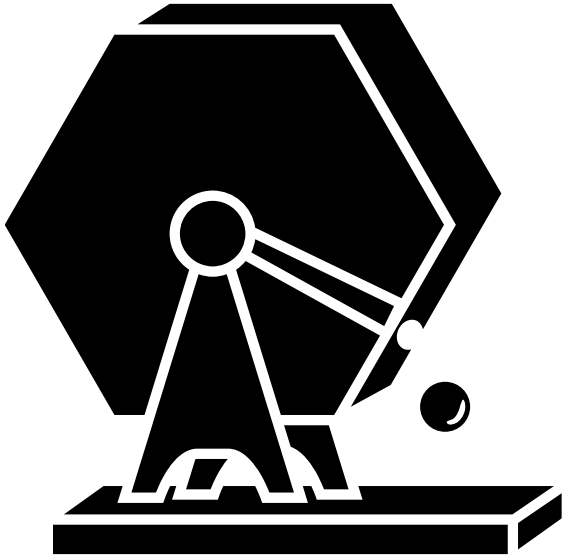
くじを戻した場合と、戻さない場合

ガラポン（下の絵のような福引き）を考えてみます。

玉が100個入っています。当りは１つで残りの99個はハズレです。以下の２つのケースを考えてみましょう。

１．１回引くたびにハズレの玉を戻して、毎回100個でくじ引きをする。
２．くじを引いても、玉を戻さずにくじ引きを続ける。

１のケースと２のケースで13番目にくじを引いた時に当りを引く確率を考えてみましょう。



くじを戻した場合と、戻さない場合（解）

1 と 2 とどちらもケースでも、13番目にくじを引いた人が当りを引く確率は $\frac{1}{100}$ です。

1 のケース：1 回引くたびにハズレの玉を戻して、毎回100個でくじ引きをする。

くじを引く事象は、毎回独立した事象と考えられます。前のくじ引きと次のくじ引きの間には、なんら関係はありません。ですから、毎回毎回100個の玉のなかから1個の当りを狙う $\frac{1}{100}$ の確率でのくじ引きが続いていきます。

2 のケース：くじを引いても、玉を戻さずにくじ引きを続ける。

くじを引く事象は、連続した事象と考えられます。それでは13番目に当りを引く確率を計算してみましょう。13番目にくじを引くということは、前の12人目まではハズレが続いていると考えると、

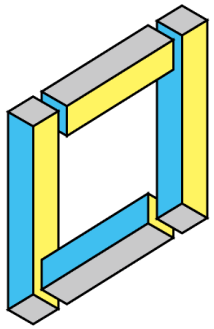
$$\frac{99}{100} \times \frac{98}{99} \times \frac{97}{98} \times \frac{96}{97} \times \frac{95}{96} \times \frac{94}{95} \times \frac{93}{94} \times \frac{92}{93} \times \frac{91}{92} \times \frac{90}{91} \times \frac{89}{90} \times \frac{88}{89} \times \frac{1}{88} = \frac{1}{100}$$

1 回目のハズレ

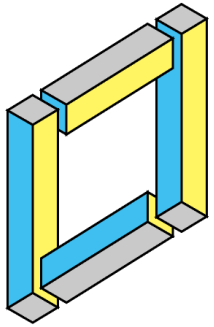
6 回目のハズレ

約分で 1 になる

13回目で当り



練習問題



袋の中の玉

袋のなかに赤い玉が３個，白い玉が３個，青い玉が４個が入っている。
同時に２個を取り出したとき，同じ色になる確率はいくらか。

袋の中の玉（解）

2つとも赤い玉が出る確率は、$\frac{3}{10} \times \frac{2}{9} = \frac{1}{15}$

2つとも白い玉が出る確率は、$\frac{3}{10} \times \frac{2}{9} = \frac{1}{15}$

2つとも青い玉が出る確率は、$\frac{4}{10} \times \frac{3}{9} = \frac{2}{15}$

それぞれの事象は同時には起きえないので、たし算で確率を求めることができる。

$$\frac{1}{15} + \frac{1}{15} + \frac{2}{15} = \frac{4}{15}$$

先に３勝する確率は？

ＡとＢの２人でチェスをしている。ＡがＢに勝つ確率が$\frac{1}{3}$であるとき、Ａが先に３勝する確率はいつくか？

先に３勝する確率は？

ＡがＢに３勝するパターンは10通りある。それぞれの確率を求めて和を出せばよい。

それぞれの試合は、独立した事象なので毎回のＡの勝率は $\frac{1}{3}$ です。

① Ａの３勝０敗が１パターン（ＡＡＡ）

$\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}=\frac{1}{27}$

② Ａの３勝１敗が３パターン（ＡＡＢＡ、ＡＢＡＡ、ＢＡＡＡ）

$\frac{1}{3}\times\frac{1}{3}\times\frac{2}{3}\times\frac{1}{3}=\frac{2}{81}$ （かけ算中の $\frac{2}{3}$ の位置は負けたパターンにより変わります。）

３パターンあるので　$\frac{2}{81}+\frac{2}{81}+\frac{2}{81}=\frac{6}{81}=\frac{2}{27}$

③ Ａの３勝２敗が６パターン（ＡＡＢＢＡ、ＡＢＡＢＡ、ＡＢＢＡＡ
ＢＡＡＢＡ、ＢＡＢＡＡ、ＢＢＡＡＡ）

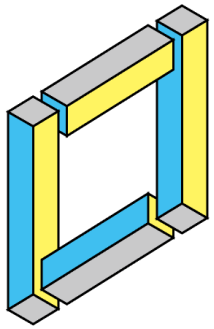
$\frac{1}{3}\times\frac{1}{3}\times\frac{2}{3}\times\frac{2}{3}\times\frac{1}{3}=\frac{4}{243}$ （かけ算中の $\frac{2}{3}$ の位置は負けたパターンにより変わります。）

６パターンあるので　$\frac{4}{243}+\frac{4}{243}+\frac{4}{243}+\frac{4}{243}+\frac{4}{243}+\frac{4}{243}=\frac{24}{243}=\frac{8}{81}$

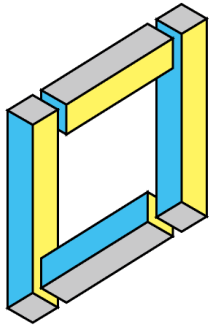
いづれも同時には起こらないので　$\frac{1}{27}+\frac{2}{27}+\frac{8}{81}=\frac{17}{81}$

データ分析以前の着眼点

確率（確率の計算）② 直感からではわからないこと



直感と計算



計算の大切さ（１）

次の３つを確率が高い順に並べてください。

※まずは計算をしないで、直感で順番を考えてから、計算をして
確認してみましょう。

① 赤玉50個、白玉50個入った袋から、１回で赤玉を取り出す。

② 赤玉90個、白玉10個入った袋から、７回連続で赤玉を取り出す。
但し取り出した玉は、毎回元の袋に戻します。

③ 赤玉10個、白玉90個入った袋から、７回連続で球を取り出し、
１回以上が赤玉である。但し取り出した玉は、毎回元の袋に戻します。

計算の大切さ（２）

実際の確率計算に反して、多くの人は②→①→③の順に確率が高いと感じるそうです。

① 赤玉50個、白玉50個入った袋から、1回で赤玉を取り出す。

→ **50%**

②　赤玉90個、白玉10個入った袋から、7回連続で赤玉を取り出す。

→ **48%**

$0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.478$

③　赤玉10個、白玉90個入った袋から、7回連続で球を取り出し、1回以上が赤玉である。

→ **52%**

$1 - (0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9) = 0.521$

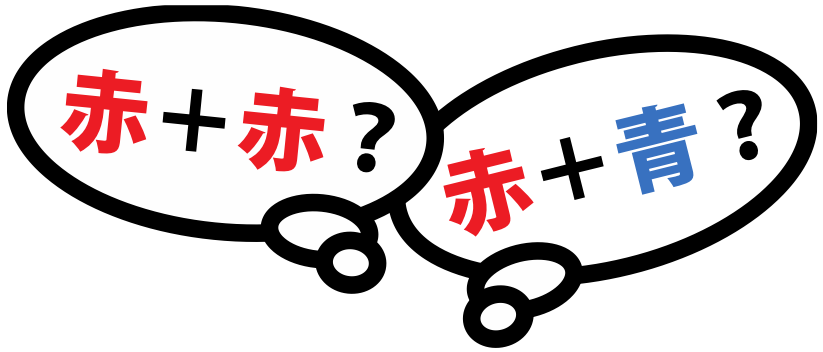
参考：ダニエル・カーネマン「ファスト＆スロー」

直感の不確定さ（１）

人間の直感がそれぞれの人生において非常に重要なモノであることは間違いないと思われますが、データ分析や統計・確率を学んでいく中で直感に頼って考察に飛びついてしまうのは非常に危険です。
ここでは、直感と計算結果が乖離している例をいくつか考えていきましょう。

簡単な例から考えていきましょう。

6面のサイコロがあります。このサイコロは数字が書かれているのではなく、色分けされています。３面が赤、２面が青、１面が黄色で塗られています。このサイコロを２個ふった時に、色の組み合わせとして一番高い確率でててくるものは何色と何色の組み合わせでしょうか？



直感の不確定さ（２）

簡単な計算で検証することができます。

直感的には、赤で塗られている面が多いので、直感的に両方とも赤になる
確率が高そうに感じられますが、それぞれの確率を計算してみましょう。

両方とも赤になる確率

$$\frac{3}{6} \times \frac{3}{6} = \frac{9}{36} = \frac{1}{4}$$

赤と青の組み合わせになる確率

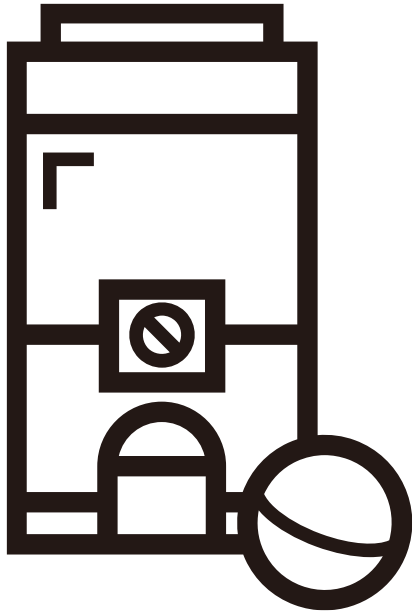
$$\frac{3}{6} \times \frac{2}{6} + \frac{2}{6} \times \frac{3}{6} = \frac{12}{36} = \frac{1}{3}$$

計算をしてみると、赤と青の組み合わせになる方が高いということが分かります。

直感と余事象（１）

５種類の景品の内、欲しいモノは１種類だけ

１回200円で引けるガチャに、５種類のフィギュアが入っている。すでに、４種類持っていて残りはあと１種。その時たまたまネット通販で、最後の1種のフィギュアが送料込み1000円でででていた。どうしてもコンプリートしたい。ガチャをやるべきか通販で買ったほうがいいか。



直感と余事象（２）

５種類の景品の内、欲しいモノは１種類だけ

欲しいフィギュアが５種類のうちの１種類。つまり確率としては$\frac{1}{5}$となります。
この余事象を考えていくとわかりやすいと思います。
$\frac{1}{5}$であたるもの＝$\frac{4}{5}$ではずれる、ということですね。
通販の1000円分を使ったとして、５回連続ではずれる確率を考えてみましょう。

$$\frac{4}{5} \times \frac{4}{5} \times \frac{4}{5} \times \frac{4}{5} \times \frac{4}{5} = \frac{1024}{3125}$$

$\frac{1}{5}$の確率と聞いた直後の直感で、５回に１回は当りが引けるように感じでしまいがちですが、余事象を考え計算すると、おおよそ$\frac{1}{3}$の確率で1000円使ってもガチャでゲットすることはできません。
もちろん運次第では、200円で手にすることもあります。

誕生日のパラドックス（１）

出会った人が自分と同じ誕生日である確率は $\frac{1}{365}$ です。
40人のクラスの中に同じ誕生日のペアがいる確率はどれくらいだろうか？



誕生日のパラドックス（２）

40人のクラスの任意の１人を選びます。任意に２人目を選んだ時に、1人目と２人目の誕生日が異なる確率は $\frac{364}{365}$ です。３人目を選んだ時に、３人目の誕生日が前の２人と異なる誕生日である確率は $\frac{363}{365}$ となります。このように4人目、５人目と全員が異なる誕生日の確率を計算していくと次のような式で計算できます。

$$\frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \frac{361}{365} \times \frac{360}{365} \times \frac{359}{365} \times \frac{358}{365} \times \frac{357}{365} \cdots \times \frac{326}{365} \fallingdotseq 0.108768$$

2人目　３人目　40人目

少なくとも１組の同じ誕生日のペアがいるということは、全員が異なる誕生日の余事象と考えられるので、

1 - 0.108768 = 0.891231

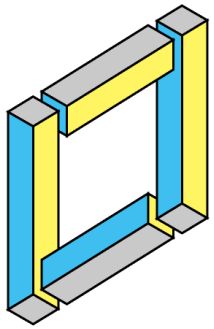
約89%の確率で40名のクラスには同じ誕生日のペアがいることがわかります。

ちなみに40人のクラスで、自分と同じ誕生日のクラスメートがいる確率は

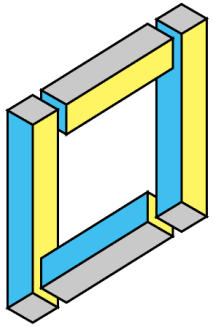
$1 - \left(\frac{364}{365}\right)^{40} \fallingdotseq 0.103932$ 約10%となります。

データ分析以前の着眼点

確率（確率の計算）③ 確率で考える・確率の利活用



期待値



期待値（１）

確率を活用していく上で「期待値」の計算は判断材料として非常に有効です。
期待値とは、ある事象で得られる値の平均値を表し、得られる可能性がある
すべての値と、それぞれが起こりえる確率の和との積求めます。

文章で読んでもイメージしづらいので、具体例を考えてみましょう。

100円で１回サイコロをふることができます。出た目に応じて下の表のよう
に景品が貰えます。サイコロの目はどの目も $\frac{1}{6}$ の確率でてます。期待値は、
どうやって求めたらいいのでしょうか。

サイコロの目	景品
1	賞品Ａ（1000円相当）
2	ハズレ
3	賞品Ｂ（300円相当）
4	ハズレ
5	賞品Ｃ（10円相当）
6	ハズレ

期待値（２）

「期待値」は、ぞれぞれが起こりえる確率と、その場合にえられる値の積の和で求めます。この場合は出たサイコロの目でもらえる景品の値と、その目が出る確率の積がその場合の値になり、すべての場合の和が期待値となります。

サイコロの目	景品	確率 $\frac{1}{6}$ をかけた金額
1	賞品Ａ（1000円相当）	約166.7円
2	ハズレ	０円
3	賞品Ｂ（300円相当）	50円
4	ハズレ	0円
5	賞品Ｃ（10円相当）	約1.7円
6	ハズレ	0円
合計		218.4円

期待値（３）

このサイコロゲームの期待値は218.4円とでました。この218.4円とは、100円払ってこのゲームに参加して得られる値の平均値となり、投資額100円に対して、218.4円というリターンが平均的に得られるということになり、かなり率のよい投資と考えることができます。

参考（期待値の公式）

期待値＝（受け取る価値×確率）の合計

一般的には、$\frac{\text{（受け取る価値×確率）の合計}}{\text{投入価値}} \times 100$の式を利用して、

百分率で表すことが多い。

期待値（４）

期待値の計算方法がわかると、次の用なケースを考える際に参考となる
数値を導き出すことができるようになる。

**それぞれの期待値を算出して
比較してみましょう。**

期待値の高いプロジェクトはどれだろうか？

１．実行予算が1000万円のプロジェクトで、20%の確率で成功する。
　　成功報酬は1200万円。

２．実行予算が500万円のプロジェクトで、70%の確率で成功する。
　　成功報酬は800万円。

３．実行予算が200万円のプロジェクトで、５%の確率で成功する。
　　成功報酬は5000万円。

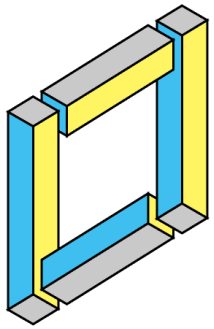
期待値（５）

１．1200万×20%=240万円
実行予算が1000万円なので、1万円あたりの投資額に対する期待値は、
240万÷1000=2400
投資1万円に対する期待額は2400円

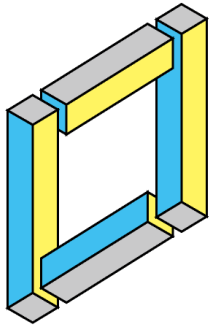
２．800万×70%=560万円
実行予算が500万円なので、1万円あたりの投資額に対する期待値は、
560万÷500=11200
投資1万円に対する期待額は1万1200円

３．5000万×5%=250万円
実行予算が200万円なので、1万円あたりの投資額に対する期待値は、
250万÷200=12500
投資1万円に対する期待額は1万2500円

3番のプロジェクトの期待値が一番高く算出されます。



条件付きの確率



条件付きの確率（１）

現実社会での確率は、今まで学んできたような単純な条件で整理できる事象は少数派です。現実は、モノゴトを確率で考えていこうとすると、事象と起こりえる結果には様々な条件が加わっているケースがほとんどです。

条件が加わった事象の確率を考えてみましょう。

ジョーカーを抜いたトランプが１セット52枚あります。この中から無作為に1枚引いた時のカードはハートでした。このカードがＡである確率はいくつか？



条件付きの確率（２）

確率を求めたい事象に対してある条件が加わった時の計算方法を考えてみます。この場合、Ａ選ぶという確率を求めるうえで、ハートのカードのなかから選ぶという条件が加わっていると考えることができます。よって確率の基本的な求め方の考えに戻って整理して考えてみると、当りは**「ハートのＡが選ばれる」**すなわち52枚のカードの1枚を選ぶことになり、確率は $\frac{1}{52}$ です。

次に場合の数を考えてみましょう。この事象はハートのカードがが選ばれた時にだけおきることなので、全体の場合から**ハートのカードが選ばれる確率**が場合の数となります。確率は $\frac{13}{52}=\frac{1}{4}$ です。

したがって、求める確率は $\frac{1}{13}$ となります。

$$\frac{\frac{1}{52}}{\frac{1}{4}}=\frac{1}{52}\div\frac{1}{4}=\frac{1}{13}$$

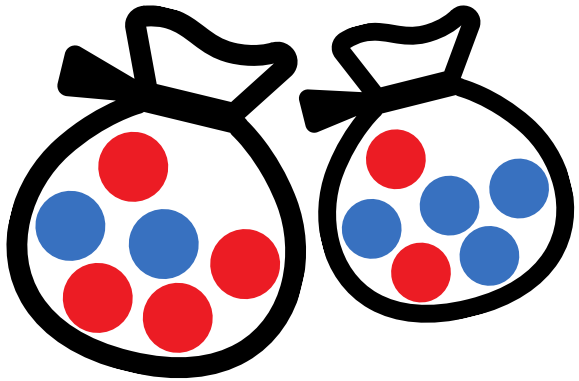
ハートのＡが選ばれる確率

ハートが選ばれる確率

条件付きの確率（３）

次の事象の確率を考えてみましょう。

ＡとＢの２つの袋があります。それぞれ６つずつボールが入っています。Ａの袋には赤いボールが４つと青いボールが２つ、Ｂの袋には青いボールが４つと赤いボールが２つ入っていることが分かっています。今、無作為にどちらかの袋からボールをひとつとり出した時に、そのボールは赤でした。このボールがＡの袋から取り出された確率を考えてみましょう。



条件付きの確率（４）

この場合、赤いボールを選ぶという確率を求めるうえで、Ａの袋から選ぶという条件が加わっていると考えることができます。よって確率の基本的な求め方の考えに戻って整理して考えてみると、当りは

「Ａの袋から選ばれる」かつ「赤いボールが選ばれる」

となります。次に場合の数を考えてみましょう。この事象は赤いボールが選ばれた時にだけおきることなので、全体の場合から**赤いボールが選ばれる確率**が場合の数となります。したがって、求める式は

$$\frac{\text{「Ａの袋から選ばれる確率」}\times\text{「赤いボールが選ばれる確率」}}{\text{「赤いボールが選ばれる確率」}} = \text{Ａの袋から赤いボールを選んだ確率}$$

となり、求める確率は $\frac{2}{3}$ となります。

Ａの袋が選ばれる確率

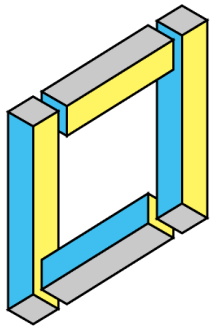
赤いボールが選ばれる確率

$$\frac{\frac{1}{2}\times\frac{4}{6}}{\left(\frac{1}{2}\times\frac{4}{6}\right)+\left(\frac{1}{2}\times\frac{2}{6}\right)} = \frac{1}{3}\div\frac{1}{2} = \frac{2}{3}$$

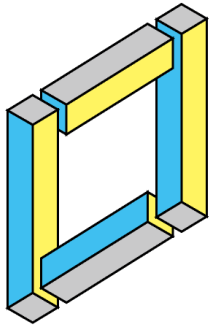
Ａの袋で赤いボールが選ばれる確率

加法定理

Ｂの袋で赤いボールが選ばれる確率

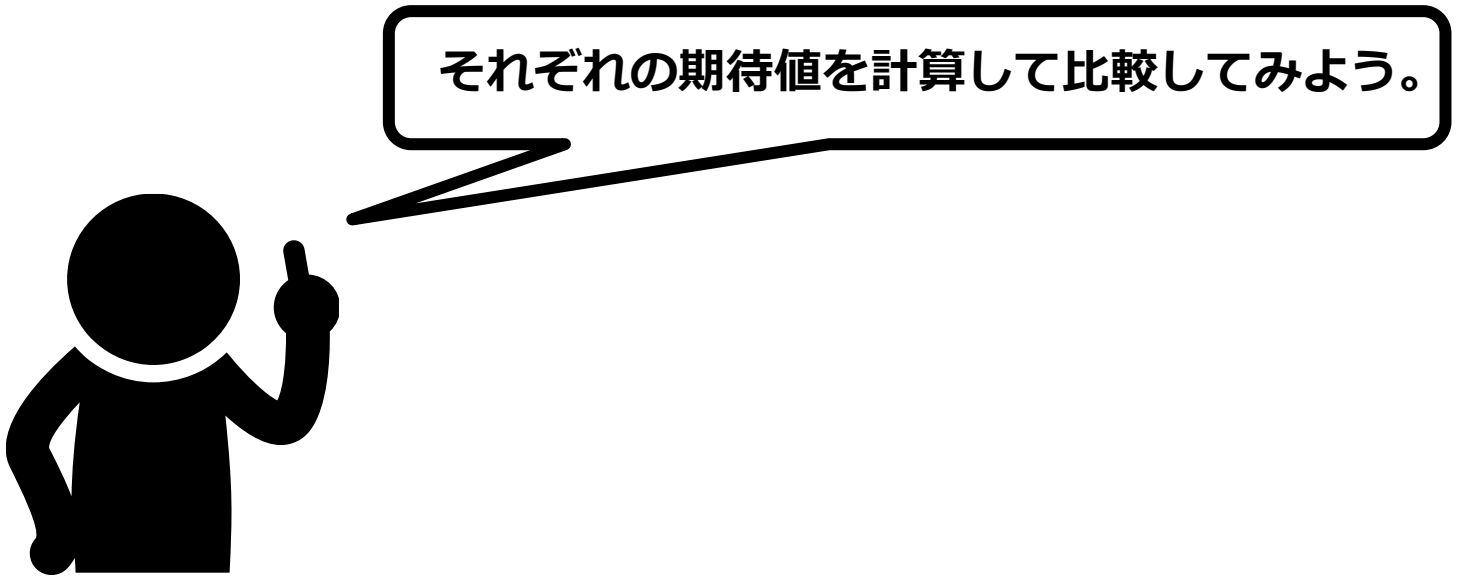


演習



演習（１）

あるところにＡ、Ｂという２つの商店があります。商店Ａでは、すべてのお買い物に対して５％引きというキャンペーンをやっています。
一方商店Bでは、それに対抗してお買い上げ20名あたりに1人無料にするというキャンペーンを始めました。
どちらがお得か考えてみましょう。



演習（解）

仮に10万円の商品を購入したとして期待値の計算をしてみます。

商店Ａ

$$10\text{万円}\times\frac{5}{100}\times\frac{100}{100}=2500\text{円}$$

5%の割引

全員が当る

商店Ｂ

$$10\text{万円}\times\frac{100}{100}\times\frac{1}{20}+10\text{万円}\times 0\times\frac{19}{20}=2500\text{円}$$

100%の割引

20人に1人が当る

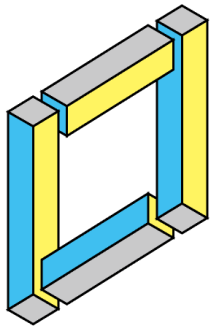
割引なし

20人に19人がはずれる

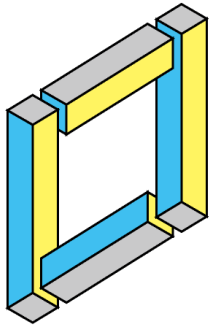
計算をしてみると、10万円の買い物に対しての期待値は同額です。お店側にとっての負担は、全く同じと考えられます。
ですが、当った時の額が商店Ｂの方が大きいので、話題性やお得感などは商店Ｂの方が上手であると思われます。

データ分析以前の着眼点

確率（場合の数）①　確率とは何か・日常にひそむ確率



確率とはなにか？



確率とは何か？

方程式や関数、図形の定理など、中学校・高校で学んできた数学には、いろいろな分野・単元がありました。そのなかで、特殊な職業に就かれている方以外で、最も日常的に聞く言葉が「確率」ではないでしょうか？

サイコロを振ったときに、1の目が出る確率は？
明日の降水確率は？
このプロジェクトの成功の確率は？

確率とは、ある事柄や行動の結果が特定の結果となる割合

つまり、「そうなるだろう」という目安

どうして、こんなことを考えるのか？

世界は、偶然でできている！

数学的確率と統計的確率

数学的確率

起きうる結果が何通りあるか計算して求める確率

統計的確率

実際に起こったことの結果を集計して求める確率

サイコロをふったとき、1が出る確率は？

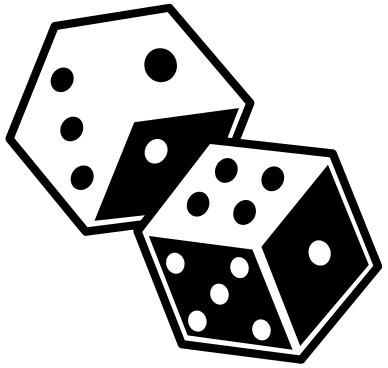
サイコロには１～６までの目があります。ふったときには、１～６の目のうち、どれかひとつに決定します。つまり、起きうる結果は１～６までの６通りで、結果は、１～６のうちのどれか一つです。
したがって、

$$1 \div 6 = \frac{1}{6}$$

決定する結果

あり得る結果のパターンの数

確率



統計的確率

今日の午後雨が降る確率は30%です。

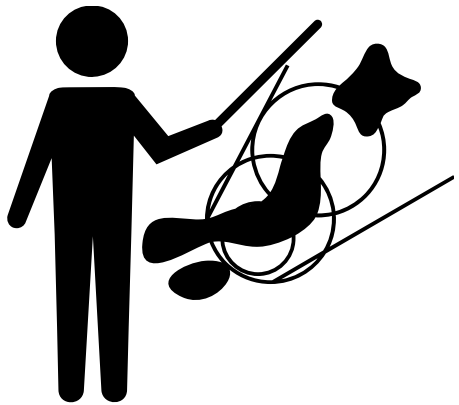
今日の午後の天気図や、その他の気象条件を過去の類似した状況に照らし合わせて、過去にどれくらいの割合で雨が降っていたかの数値を確率としてだしたものです。
したがって、計算の意味は

$$30 \div 100 = \frac{3}{10}$$

確率

過去の同様な気象条件の時に、雨が降った回数

調べた過去の同様な気象条件の件数



確率の求め方

場合の数とあたりの数

$$\text{確率} = \frac{\text{当てはまる数}}{\text{場合の数}}$$

ここの算出の仕方の例を
見ていきましょう。

サイコロをふって、1が出る確率

$$1 \div 6 = \frac{1}{6}$$

サイコロをふって、偶数が出る確率

2,4,6のどれかが
出ればいいから、
当てはまる数は3

$$3 \div 6 = \frac{3}{6} = \frac{1}{2}$$

確率の計算では、場合の数を正しく算出することが非常に重要!

大数の法則

たとえば、今ここでサイコロを６回振ったとしましょう。はたして、
１の目は何回出るでしょうか？

数学的確率から考えていけば、１回は1の目が出るように思われますが、実際はそうはならないケースの方が多いのではないでしょうか。これは、サイコロの出目は、偶然によって偏るという事から起こります。

しかし、サイコロを振っていく回数を増やしていくと、どんどん１の目が出る確率は$\frac{1}{6}$に近づいていきます。60回振った場合では、8回であったり13回であったものが、600回、6000回、60000回と続けていくことにより、どんどん$\frac{1}{6}$に近づいていきます。

このように、試行を繰り返していくことで、統計的確率が数学的確率に近づいていくことを、**「大数の法則」**といいます。

確率の豆知識 ①

年末ジャンボ宝くじで１等に当選する確率と自動車事故で死亡する確率

年末ジャンボの１等にあたる確率を考えてみましょう。

年末ジャンボの販売枚数は4億4千枚万です。１等7億円は22本あります。１等の当る確率は次の計算式で求められます。

数学的確率

$$22 \div 440,000,000 = \frac{1}{20,000,000}$$

2000万分の１！

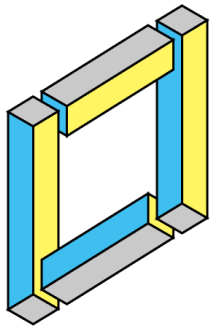
今度は、自動車事故で死亡する確率を考えてみましょう。

日本の人口は１億2580万人です。警察庁の発表によると令和元年の交通事故による死者数は3215人です。（1年間の確率です。）

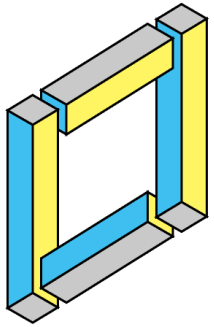
統計的確率

$$3,215 \div 125,000,000 = \frac{1}{40,000}$$

４万分の１！



練習問題



２つのサイコロを投げた時の確率について

大小２つのサイコロを振った時、次の確率を求めなさい。

１．２つのサイコロの出目が同じ確率は？

２．大きいサイコロの出目が、小さいサイコロの出目よりも大きい確率は？

３．２つのサイコロの出目の和が、７になる場合と７になる場合では、
どちらの方が確率が高いか？

２つのサイコロを投げた時の確率について（解１）

組み合わせ表を使って、確率を考える。組み合わせ表は、すべての場合の数を考えることができるので、確率を考える時に有効な手段のひとつです。

１．２つのサイコロの出目が同じ確率は？

小＼大	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

すべての組み合わせのパターンは36通りあります。その中で、出目が同じパターンは６通りです。したがって確率は、

$$\frac{6}{36} = \frac{1}{6}$$

当てはまる数

場合の数

２つのサイコロを投げた時の確率について（解２）

２．大きいサイコロの出目が、小さいサイコロの出目よりも大きい確率は？

小＼大	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

オレンジでマークした部分が辺りのパターンとなります。

$$\frac{15}{36} = \frac{5}{12}$$

当てはまる数

場合の数

２つのサイコロを投げた時の確率について（解３）

３．２つのサイコロの出目の和が、６になる場合と７になる場合では、
どちらの方が確率が高いか？

小＼大	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

和が６になる確率　$\frac{5}{36}$

和が７になる確率　$\frac{6}{36} = \frac{1}{6}$

よって、和が7になる確率の方が高い

３つのサイコロを投げた時の確率について

３つのサイコロを振った時、出目の和が9になる場合と10になる場合では、
どちらの確率が高いか考えてみましょう。

３つのサイコロを投げた時の確率について（解）

サイコロを３つ振った時に出目の和が９になるパターンは、(1,2,6),(1,3,5),(1,4,4),(2,2,5),(2,3,4),(3,3,3)の6パターン

サイコロを３つ振った時に出目の和が10になるパターンは、(1,3,6),(1,4,5),(2,2,6),(2,3,5),(2,4,4),(3,3,4)の6パターン

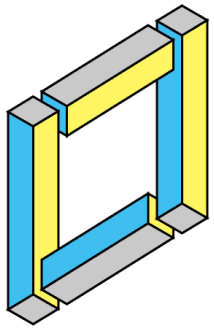
どちらも、6パターンなので、同じ確率のように考えてしまいそうですが、大小２つのサイコロの時のように組み合わせ表を考えてみましょう。大きな表になるので、現実的に書いてみるのは面倒ですが、大中小の３つのサイコロの組み合わせを考えていくと次のように考えられます。

(1,2,6)のパターン
(1,2,6), (1,6,2),(2,1,6),(2,6,1),(6,1,2),(6,2,1)

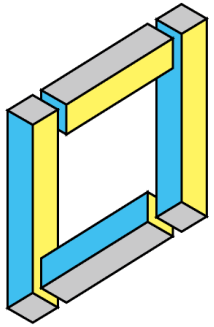
このように、和が9になるパターンと、10になるパターンのそれぞれのサイコロの組み合わせを順番をつけて考えていくと、和が9になるパターンは25通り、和が10になるパターンは27通りあります。
したがって、3つのサイコロを振った時に和が10になる確率の方が和が9になる確率より高いと言えます。

データ分析以前の着眼点

確率（場合の数）②　いかにして数えるか・場合の数の基本



場合の数・当りの数の求め方



ドリンクバーでのソフトドリンクの選択と、好きなコーヒーの選択

８種類のドリンクがあるドリンクバーで、１回好きなものを飲める。（混ぜるのはなし！）

このときの場合の数は、8ですよね。

それでは、好きなコーヒーを選んだ場合の場合の数を次の条件から考えてみましょう。

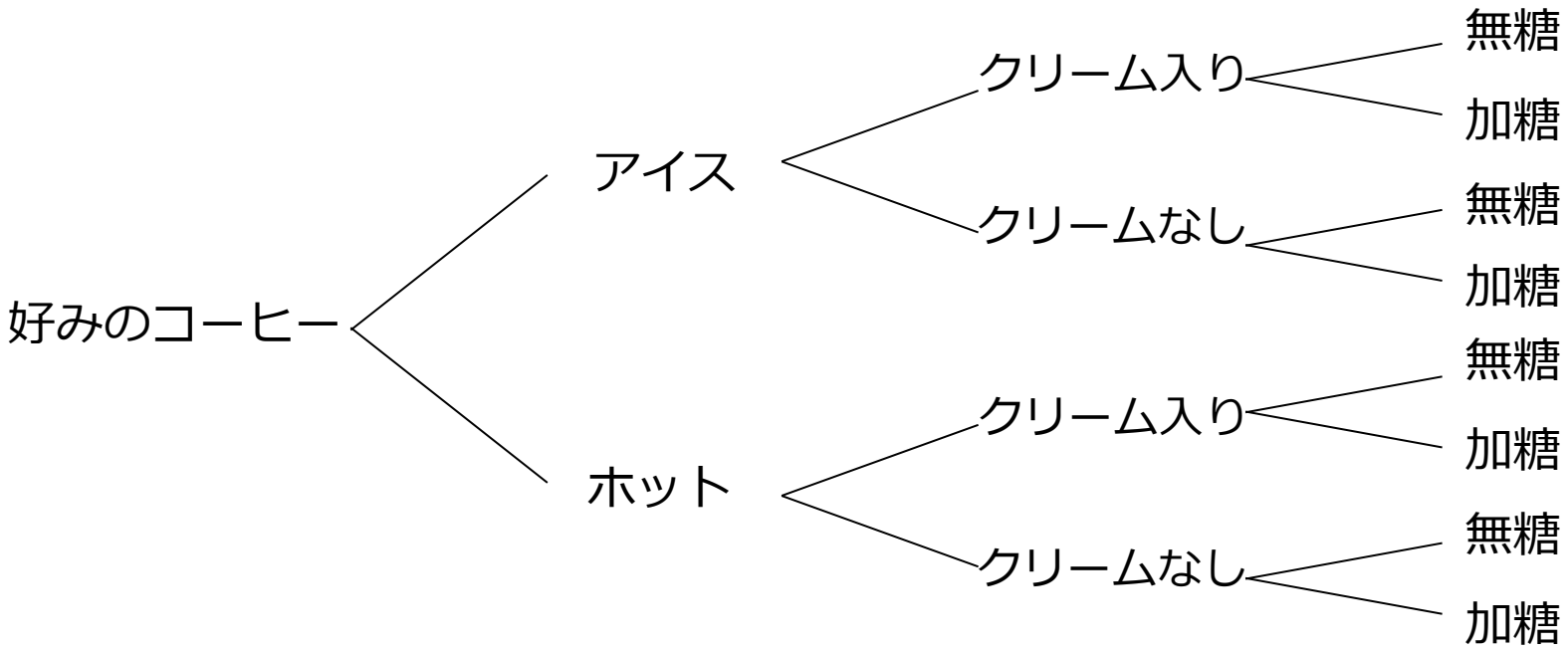
選択出来るチョイス

1．アイスかホット
2．クリームは入れるか入れないか
3．砂糖は入れるか入れないか



樹形図で整理してみましょう。

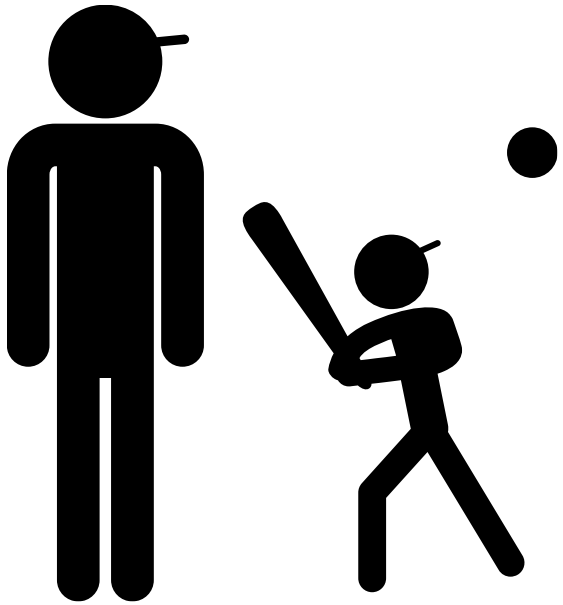
樹形図を書くと、分岐とパターンがわかりやすくなり、場合の数を求めるときに、間違いが起こりにくくなります。



2つの選択肢で３階分岐する、つまり $2^3 = 8$ の場合の数となります。

階乗を使って計算する１

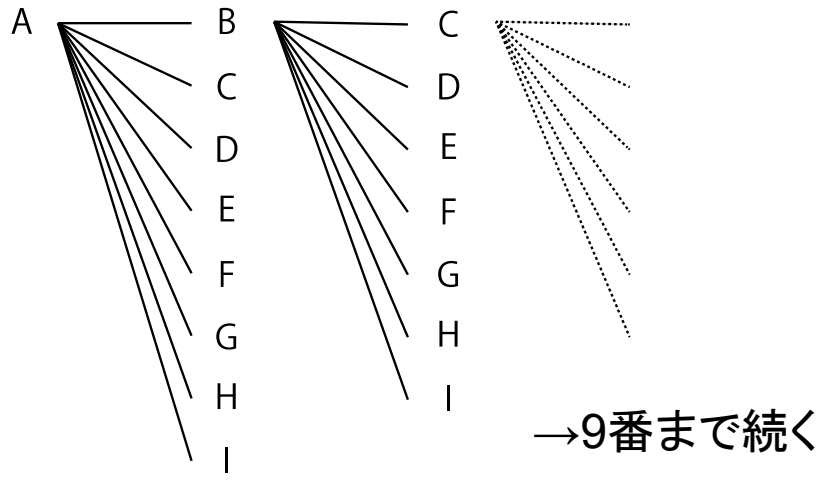
少年野球のチームでやっとメンバーが９人揃った。各ポジションはコーチと選手の話し合いで決まったが、打順が決まらない。そこで、コーチは、打順１番から９番まで、すべてのパターンを試してみてから決めることを提案した。打順は何パターンあるのだろうか？



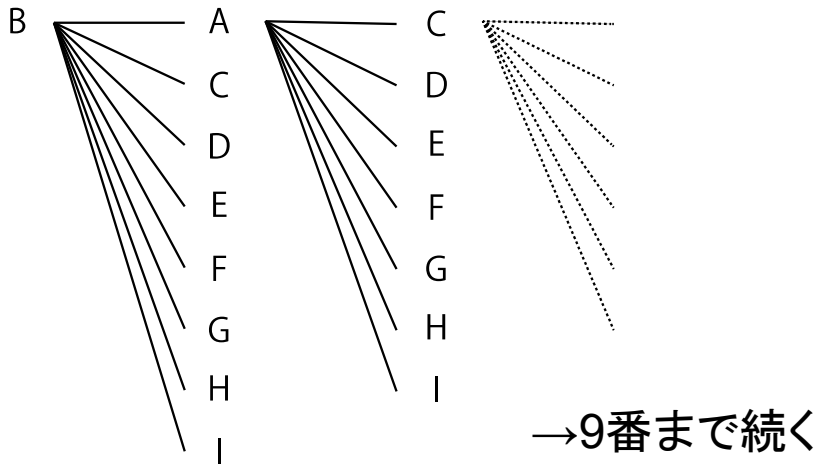
階乗を使って計算する２

ＡからＩまでの9人がいる。打順1番に入れる可能性は９人。２番に入れるのは、１番に入らなかった８人、３番に入るのは１番２番に入らなかった７人と、言うカタチで順番を考えていくことができる。

Aを1番とした時



Bを1番とした時



階乗を使って計算する３

１番に入ることの出来る選手は９人。２番に入ることの出来る選手は１番に入らなかった8人。３番に入ることのできる選手は、１番と２番に入らなかった７人、というようにかけ算で求めることができます。

$$9\times8\times7\times6\times5\times4\times3\times2\times1=362,880$$

この式を階乗記号「！」を使って表すことができます。

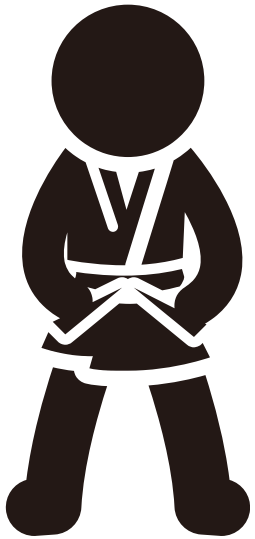
$$9!=362,880$$

仮に１日１パターンの打順を試してみて、毎日試合をしたとして全部のパターンを試すまでに約1000年かかります。あまり現実的な提案とは言えません。

順列を考える１

今度は、少年野球ではなくある中学の柔道部の場合を考えてみましょう。

柔道部員が８名いる。大会に出るために5人の選手を決めなければならない。
先鋒、次鋒、中堅、副将、大将のパターンは何通り考えられるか？



順列を考える２

野球の打順の時のように、順番が関係ある並び方のパターンを考えますが、
今回は、５人選んだ後の３人の順番はパターンには関係ありません。
ですので、次のような式で計算することができます。

$$8\times7\times6\times5\times4\times3\times2\times1=6,720$$

ここの順番は関係ないので計算に含めない。

参考（順列の公式）

異なるｎ個のものからｒ個選んだときの順列＝${}_nP_r=\frac{n!}{(n-r)!}$

この問題の場合、 ${}_8P_5=\frac{8\times7\times6\times5\times4\times3\times2\times1}{3\times2\times1}=6,720$

組み合わせを考える１

次は、ある高校のバスケットボール部の場合を考えてみましょう。

バスケットボール部員が８名いる。レギュラー5人を選んだ時のパターンは何通りあるか？ただし、このバスケットボールチームにはポジションの考えがないものとする。



組み合わせを考える２

野球や柔道の時と違って、レギュラー５人選考したときの順番は関係ありません。
その時には、計算方法はどう変わるのでしょうか。選手の８名をＡからＨとします。
順列の時のように５人選考したときのことを考えてみましょう。
仮に、Ａ、Ｂ、Ｃ、Ｄ、Ｅの５人が選ばれたとします。
このときに、順列では別のパターンとして考えていた、Ｂ、Ｄ、Ａ、Ｅ、Ｃや、Ｅ、Ｂ、Ｄ、Ｅ、Ａなどは、同じメンバーで構成されているので、重複していると考えられます。ですから、順列で求めたパターンから重複している部分を取り除いて行けばいいわけです。
たとえばこの場合ならば、８人から選んだ５人の順列のパターン数を重複する５人の順列（同じメンバーで順番が違う）で割ってあげればいいわけです。

$$(8\times7\times6\times5\times4)\div(5\times4\times3\times2\times1)=56$$

参考（組み合わせの公式）

異なるｎ個のものからｒ個選んだときの組み合わせ＝$_nC_r=\frac{_nP_r}{r!}$

この問題の場合、$_8C_5=\frac{8\times7\times6\times5\times4}{5\times4\times3\times2\times1}=56$

順列と組み合わせのまとめ

順列 選んだものの順番や順序が関係ある場合。

異なるｎ個のものからｒ個選んだときの順列＝$_{n}P_{r}=\frac{n!}{(n-r)!}$

・10個の中から、４つ選んだ場合

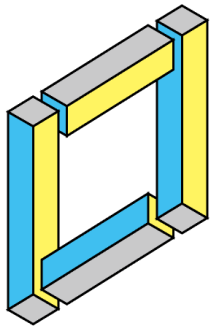
$$_{10}P_{4}=\frac{10!}{(10-4)!}=\frac{10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{6\times5\times4\times3\times2\times1}=5{,}040$$

組み合わせ 選んだものの順番や順序が関係がない場合。

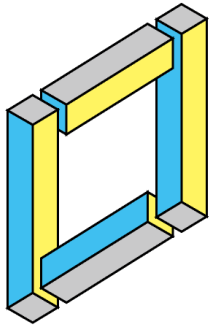
異なるｎ個のものからｒ個選んだときの組み合わせ＝ $_{n}C_{r}=\frac{{}_{n}P_{r}}{r!}$

・10個の中から、４つ選んだ場合

$$_{10}C_{4}=\frac{{}_{n}P_{r}}{4!}=\frac{(10\times9\times8\times7\times6\times5\times4\times3\times2\times1)\div(6\times5\times4\times3\times2\times1)}{4\times3\times2\times1}=210$$



練習問題



順列と組み合わせの練習問題

１． ５人の陸上部員がいる。100メートル走の選考会で、1位になった選手は奨学金と大学への推薦がもらえる。2位になった選手は、奨学金はもらえないが大学への推薦がもらえる。このとき、１位と２位になった選手の組み合わせは何通り考えられるか。

２． ５人の陸上部員がいる。100メートル走の選考会で、1位と２位になった選手はオリンピック候補生として強化合宿に招待される。このとき、１位と２位になった選手の組み合わせは何通り考えられるか。



順列と組み合わせの練習問題（解）

1．５人から２人を選んだ順列となる。
したがって、${}_5P_2$となり、
計算は (5 × 4 ×3 ×2 ×1)÷（ 3 ×2 ×1)=20

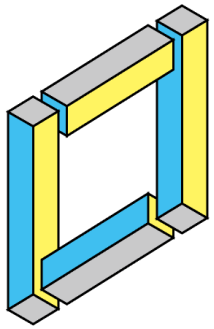
20通りのパターンが考えられる。

2．５人から２人を選んだ組み合わせとなる。
したがって、${}_5C_2$となり、
計算は (5 × 4)÷（ 2 ×1)=10

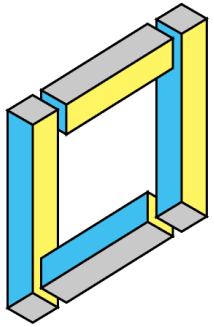
10通りのパターンが考えられる。

データ分析以前の着眼点

確率（場合の数）③　複雑なものを簡単に数える考え方



余事象



余事象とは１

こんな状況を考えてみましょう。

くじ引きをしました。10本あるくじの中で当りは3本です。

３回くじを引きました。少なくとも１回は当りを引く確率は？
但し、引いたくじは、そのつどもとに戻すものとします。



余事象とは２

少なくとも１回は当りという事を考えてみましょう。

１回引いた時の当りの確率は、場合の数が10で、当りが３ということから
$\frac{3}{10}$ということがわかります。

３回くじを引くということから、次のパターンが考えられます。

- ３回とも当り。（当り,当り,当り）
- ２回当り。（当り,当り,ハズレ), （当り,ハズレ,当り), （ハズレ,当り,当り）
- １回当り。（当り,ハズレ,ハズレ), （ハズレ,当り,ハズレ), （ハズレ,ハズレ,当り）

余事象とは３

それぞれの場合を求めてみましょう。

- ３回当り。　(当り,当り,当り)　$\frac{3}{10}\times\frac{3}{10}\times\frac{3}{10}=\frac{27}{1000}$
- ２回当り。　(当り,当り,ハズレ)　$\frac{3}{10}\times\frac{3}{10}\times\frac{7}{10}=\frac{63}{1000}$
  (当り,ハズレ,当り)　$\frac{3}{10}\times\frac{7}{10}\times\frac{3}{10}=\frac{63}{1000}$
  (ハズレ,当り,当り)　$\frac{7}{10}\times\frac{3}{10}\times\frac{3}{10}=\frac{63}{1000}$
- １回当り。　(当り,ハズレ,ハズレ)　$\frac{3}{10}\times\frac{7}{10}\times\frac{7}{10}=\frac{147}{1000}$
  (ハズレ,当り,ハズレ)　$\frac{7}{10}\times\frac{3}{10}\times\frac{7}{10}=\frac{147}{1000}$
  (ハズレ,ハズレ,当り)　$\frac{7}{10}\times\frac{7}{10}\times\frac{3}{10}=\frac{147}{1000}$

なぜたし算なのかは、
次回のセクションで
詳しく学びます。

少なくとも１回は当る確率は、このすべての確率の和になる。

$$\frac{27}{1000}+\frac{63}{1000}+\frac{63}{1000}+\frac{63}{1000}+\frac{147}{1000}+\frac{147}{1000}+\frac{147}{1000}+\frac{147}{1000}=\frac{657}{1000}$$

余事象とは４

少し考え方を変えてみましょう。

少なくとも１回は当りを引くというパターンの逆を考えてみましょう。

３回ともハズレを引く

- ３回ハズレの確率は、$\frac{7}{10} \times \frac{7}{10} \times \frac{7}{10} = \frac{343}{1000}$

3回ともハズレでなかった時が、少なくとも１回は当った時と考えると、

$1 - \frac{343}{1000} = \frac{657}{1000}$　と、考えることが出来る。

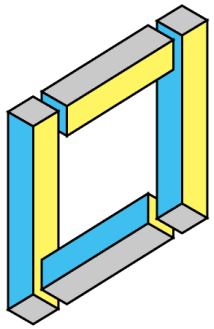
すべての場合の数
すなわち100%= 1

余事象

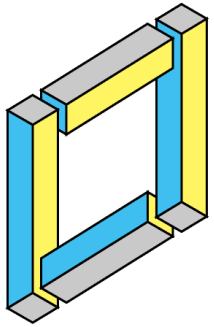
ある事象Ａが起きる確率をaとすると

１－a=事象Ａが起きない確率

と、求めることができる。



順列・組み合わせの応用



同じものを含む順列１

スマホのロックを解除する。

スマホのロック解除に4桁の数字を使ったPINコードを使用している。忘れないように、自分の誕生日である11月25日の4つの数字を並べ替えて使用すると、何通りのパターンがあるだろうか？



同じものを含む順列２

仮に4つの数字がすべて違っていたら、階乗の計算でパターン数は求められます。

$$4!=4\times3\times2\times1=24$$

24通りとなります。
ですが実際は、
(1,1,2,5),(1,1,5,2),(1,2,1,5),(1,2,5,1),(1,5,1,2),(1,5,2,1),
(2,1,5,1),(2,1,1,5),(2,5,1,1),(5,1,1,2),(5,1,2,1),(5,2,1,1)
の12通りしかありません。

これは１が２つある、つまり区別のつかない重複したモノが含まれる順列として考える必要があります。

同じものを含む順列３

重複している数字「1」を、仮に「1a」「1b」と区別したら前のページのように階乗の計算でも問題ありませんが、実際に「1a」「1b」は同じものなので、重複して数えていることになります。

(1a,1b,2,5)
(1b,1a,2,5)
→ (1,1,2,5)を重複して数えている。

よって、正しい数え方は４つの順列数を求めた後に、重複している２つの順列数で割ることで求められます。

$$4!\div 2!=\frac{4\times 3\times 2\times 1}{2\times 1}=12$$

全体の順列

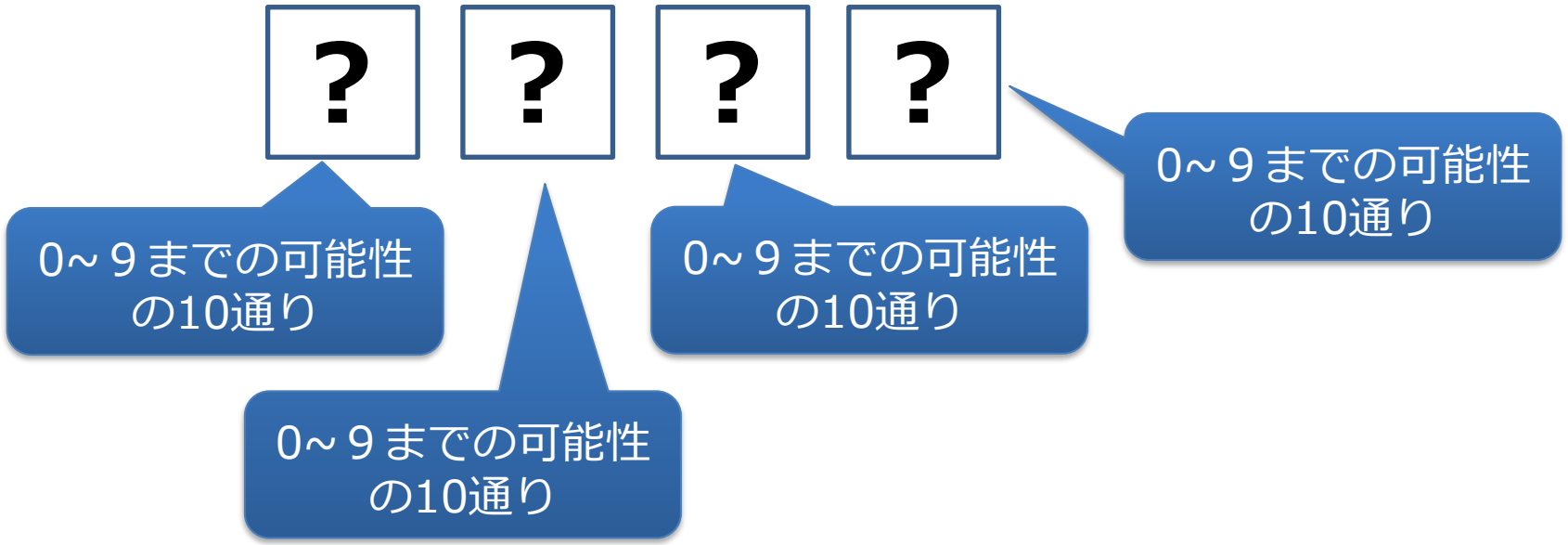
重複しているモノの順列

スマホのロックを解除する。その２

スマホのロック解除に4桁の数字を使ったPINコードを使用している。PINコードが全く思い出せなくなった時、その組み合わせは何通りあるだろうか？

重複順列２

それぞれの場所に、何通りの可能性があるかを求めていけばよい。

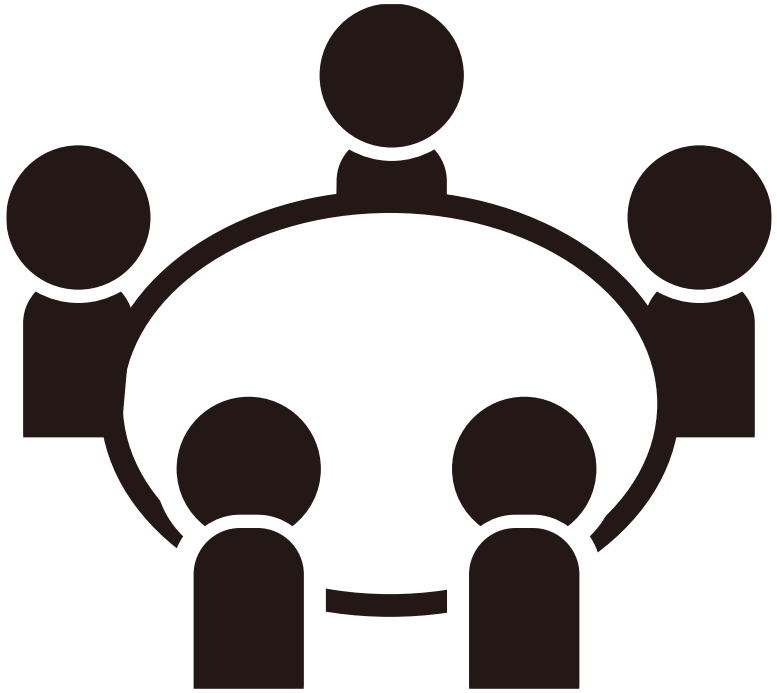


$10^4=10\times 10\times 10\times 10=10{,}000$

円順列 1

５人家族の中華料理店での話

おばあちゃん、お父さん、お母さん、長男である自分、妹の５人で中華料理を食べに行った時のことを考えてみましょう。
丸いテーブールに５人が着席した時の並び方のパターンはいくるあるだろうか？



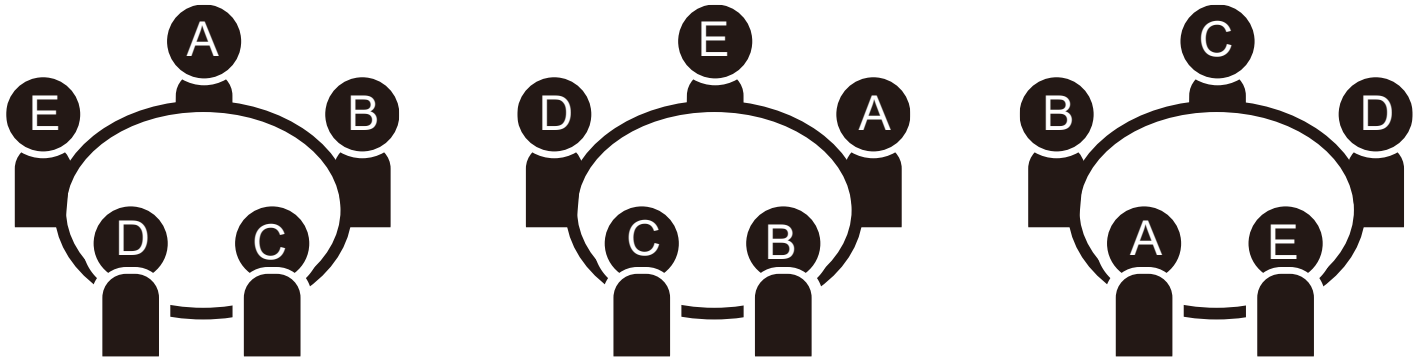
円順列２

自分を基準に考えてみよう

例えば、５人家族で１列に並んだ場合の順列は、５の階乗で求められました。

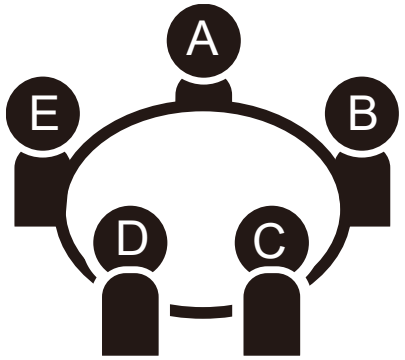
$$5!=5\times4\times3\times2\times1=120$$

120通りの並び方があります。
しかし、円いテーブルに座った場合、位置関係に気をつける必要があります。
例えば、下の３つパターンは基準位置が異なるだけで並び方は同じです。



円順列３

このような円順列の場合、一つの動かない基準を考え、残りの組み合わせを考えることによって、全体の順列数を求めることが出来ます。



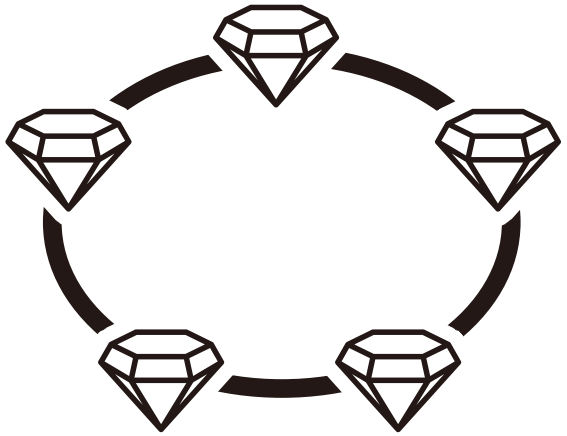
仮にAを固定して考えると、Aを除く４人の順列を考えれば並び方のパターン数は求められます。
したがって計算は、

$$(5-1)!=4\times3\times2\times1=24$$

24通りになります。

円順列の特別な例

ダイヤモンド、ルビー、サファイア、エメラルド、真珠の５つの宝石を並べてブレスレットを作ります。並べ方は何パターンあるだろうか？



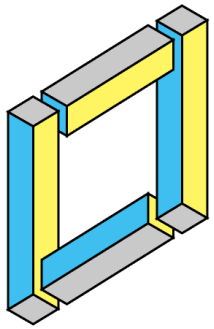
裏と表がない円順列を、数珠順列と呼びます。この数珠順列では、円順列で求めたパターン数を裏と表の２つの重複と考えることによって、求める事が出来ます。

$$\frac{(5-1)!}{2}=\frac{4\times3\times2\times1}{2}=6$$

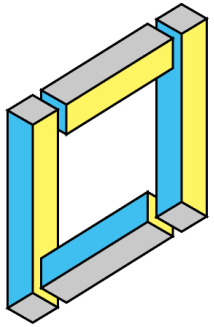
６通りの組み合わせが考えられます。

データ分析以前の着眼点

実数と割合の使い分け



実数と割合



実数と割合（１）

次の表を見てみましょう。この表は、Ａ市、Ｂ町、Ｃ村の人口と面積とコンビニの店舗数をまとめたものです。

	Ａ市	Ｂ町	Ｃ村
人口	500,000	50,000	5,000
面積	75㎢	25㎢	50㎢
コンビニの店舗数	50店	3店	6店

１．どの自治体にコンビニがたくさんあるのか考えてみましょう。

２．もし自分がコンビニのオーナーになるならば、どこで新規営業を始めるか考えてみましょう。

実数と割合（２）

次の表を見てみましょう。この表は、Ａ市、Ｂ町、Ｃ村の人口と面積とコンビニの店舗数をまとめたものです。

	Ａ市	Ｂ町	Ｃ村
人口	500,000	50,000	5,000
面積	75㎢	25㎢	50㎢
コンビニの店舗数	50店	3店	6店

１．どの自治体にコンビニがたくさんあるのか考えてみましょう。
店舗数で比較するとA市のコンビニが一番多いですが、
人口に対する割合や面積に対する割合だと多い都市が変わります。

２．もし自分がコンビニのオーナーになるならば、どこで新規営業を始めるか考えてみましょう。
店舗からのある距離以内に住んでいる人口の多さで選ぶ方法も、
周りに店舗が少ない場所を選ぶ方法もあります。

標本誤差

平均点が60点のテストがあります。
そのテストで自分が100点をとった時の影響について考えてみましょう。

１．テストの受験者が10人だった場合。

平均点は（60×10＋100）/ 11 = 63.6点で、3点以上上がります。

２．テストの受験者が10000人だった場合。

平均点は（60×10＋10000）/ 10001 = 60点で、変わりません。

サンプルが少ない時は、一つ一つのデータの影響が大きくなります。
反対に、サンプルが多い時は、特異なデータの影響が小さくなります。

サンプルから得られた値が、母集団とどれくらい一致しているか（標本誤差）を考えることが重要です。

条件付き割合（１）

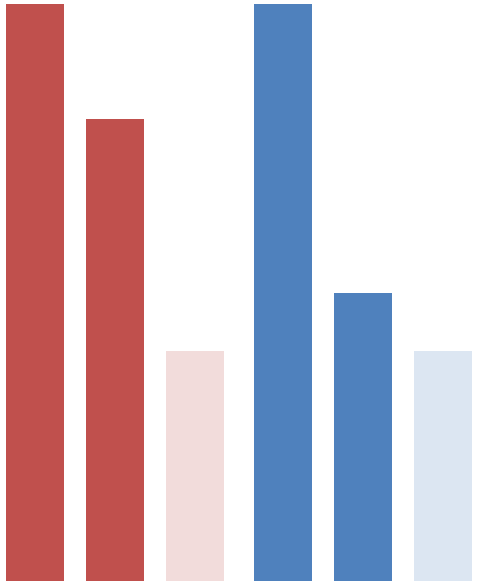
ある新規スポーツ用品ブランドの認知度を、競技場でアンケート調査を行った時、1000人のサンプルの80%が「知っている」と答え、「知っている」と答えたサンプルの50%が「良いイメージを持っている」と答えた。

ある新規スポーツ用品ブランドの認知度を、郊外のショッピングモールでアンケート調査を行った時、10000人のサンプルの50%が「知っている」と答え、知っていると答えたサンプルの80%が「良いイメージを持っている」と答えた。

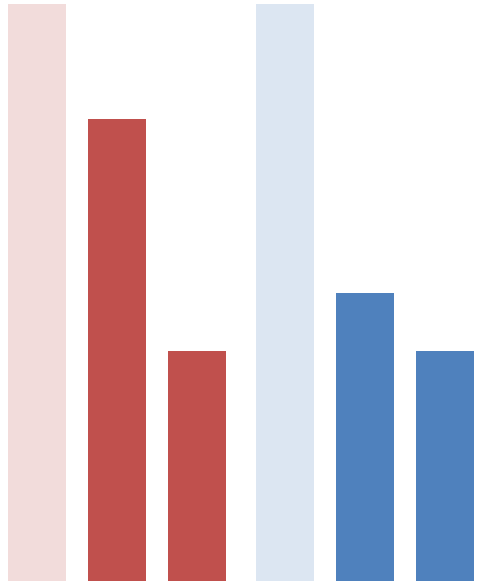
どちらも、あるブランドを「知っていて良いイメージを持っている」人のサンプル全体に対する割合は等しく40%ですが、、、

条件付き割合（２）

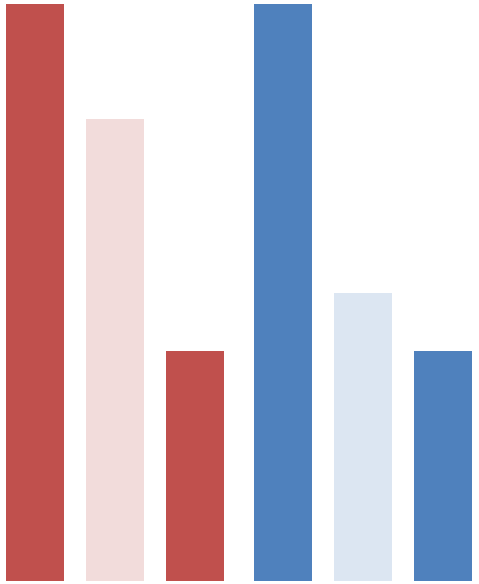
割合は、基準をとる場所（条件）により変化します。



サンプル全体を基準に
知っている割合を考える



知っている人を基準に
好印象の割合を考える



サンプル全体を基準に
好印象の割合を考える

※赤い棒 … 競技場でのアンケート調査
青い棒 … ショッピングモールでのアンケート調査
一番左をサンプル全体として1にした時の、「知っている割合（中央）」と
「知っていて良いイメージを持っている割合（右）」

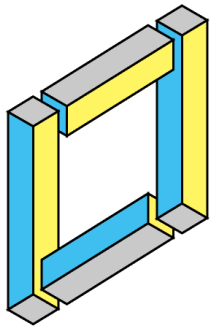
実数分析と比率分析

実数分析

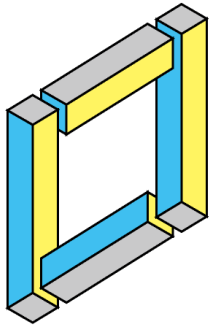
- 実際の数値を用いて分析をします。
- 直感的にわかりやすく、「量」の比較を行えます。
- 具体的には…
  - 東京駅からの所要時間で交通の便を比較する。
  - SNSの投稿数から話題性を分析する。

比率分析

- 数値の比率を求めて分析をします。
- 効率性や有効性などの、「質」の比較を行えます。
- 具体的には…
  - 所要時間に対する費用で交通手段を比較する。
  - 日本円と外貨の為替レートを見て景気を分析する。



ワーク



ワーク（1）

人口100万人の地域で、90%の住民が
ある病気の予防接種を受けました。

その病気に罹った4万人のうち、
予防接種を受けた人は3万人、
予防接種を受けていない人は1万人でした。

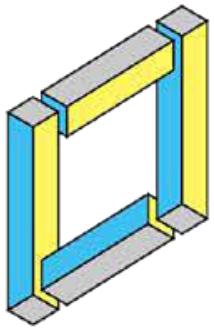
予防接種を受けた効果はあるのでしょうか？



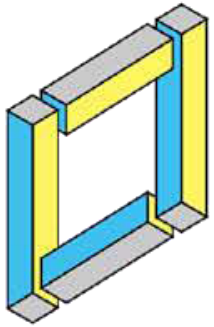


データ分析以前の着眼点

データの偏り ① なぜデータは偏るのか・代表的なデータの偏り方



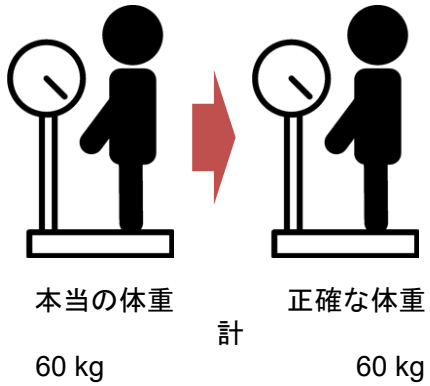
バイアスとは

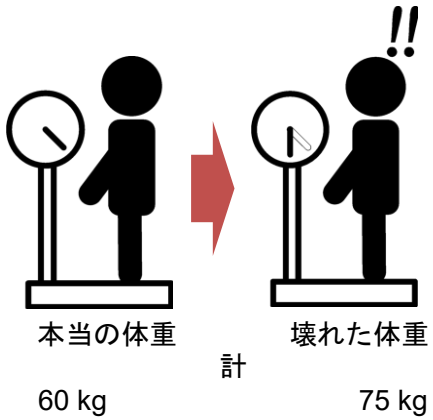


イントロダクション（1）

会社内で体重測定のデータを集めるとします。
次のどちらの調査結果の方が信頼できるでしょうか。

① 正確な体重計を使って
　　体重を計測する

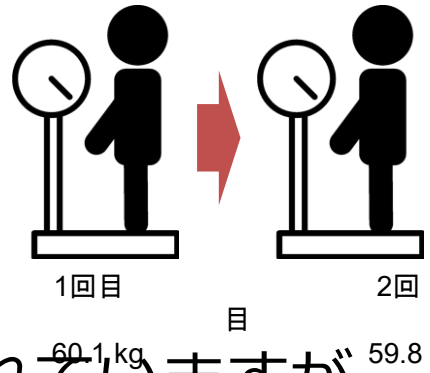




イントロダクション（２）

もちろん①の方が正確だと考えられます。
②のようにずれが起きる偶然ではない原因があって、
対処が可能な誤差を**系統誤差**と言います。

では実際の体重が60.0 kgの人が①の正確な体重計で測定を行い
1回目は60.1 kg、2回目は59.8 kg
という結果になったとします。



これは現実の60.0 kgからはわずかにずれていますが、
偶然起こったずれなので、この誤差は防ぎようがありません。
このような誤差は**偶然誤差**と言います。

バイアスとは？

バイアス：集めたデータと実際との間の偶然的でないずれ（系統誤差）を引き起こす要因のこと。

常に15kgずれている体重計のように、偶然的でない原因によって結果が現実と異なっている場合、「**バイアスがある**」と言います。

体重を測るごとに僅かにずれてしまうといった、完全に偶然起こってしまい防ぐことができないずれ（偶然誤差）があっても、「バイアスがある」とは言いません。

バイアスの種類

バイアスは発生原因によって以下の2種類に分類されます。

選択バイアス

母集団の代表である標本の選び方により生じる。

情報バイアス

対象者から情報を得る際に生じる。

バイアスの例（1）

次の例はどちらも現実とは異なる調査結果となっています。
それぞれのずれの原因は選択バイアス、情報バイアスどちらでしょうか。

① 日本人の平均身長の調査において、データを小学校３校から集めた結果、日本人の平均身長は130 cmだとわかった。



② 社内旅行の満足度調査において、記名式でアンケートを行ったところ、「非常に満足」という回答が9割を占めた。なお、社員たちは企画者がアンケートを見ると事前に知らされていた。



バイアスの例（2）

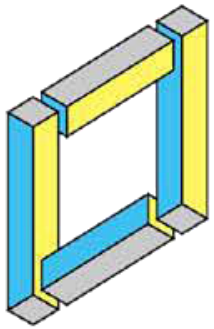
次の例はどちらも現実とは異なる調査結果となっています。
それぞれのずれの原因は選択バイアス、情報バイアスどちらでしょうか。

① 選択バイアス

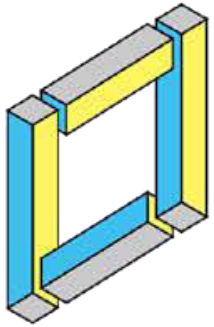
日本人全体（母集団）の平均身長を知りたいのに、標本を小学生としているためずれが生じています。つまり、選んだ標本が母集団を代表できておらず、標本の選び方が不適切だと言えます。

② 情報バイアス

企画者に回答を見られるとわかっていたため、悪いことは書きにくいという気持ちが働いて、本心とは無関係に「非常に満足」と回答する人が増えたと考えられます。つまり、情報収集の段階で現実とのずれが生じています。



選択バイアス



選択バイアスの種類

選択バイアスにはたとえば以下のようなものがあります。

- **志願者バイアス（＝自己選択バイアス）**
- **未回答者バイアス**
- **脱落バイアス**

志願者バイアス

志願者バイアス（volunteer bias）

多くの場合、調査に自ら志願して参加する人には特有の傾向があるため、標本が母集団の代表とならずに起こるバイアス。



Example

定期検診への参加が死亡率に影響を及ぼすのかを調査したいとします。

検診への参加者は非参加者よりも健康意識が高く、より健康的な生活を送っている可能性があります。

そのため、このような場合は検診の非参加者よりも参加者の方が死亡率が低いとわかったとしても、検診自体が死亡率を下げたとは言い切れません。

未回答者バイアス

未回答者バイアス（non-respondent bias）

調査に回答してくれる人と回答してくれない人とでは、傾向に差が出ることが多いため起こるバイアス。

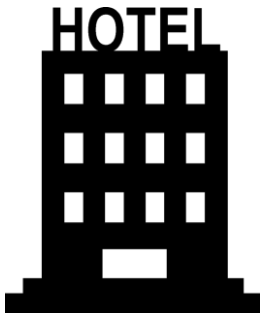
Example

ある地域への訪問者の宿泊の目的の調査を行うとします。

その地域内の全ホテルにて宿泊者に対してアンケート用紙を配布し、ホテル内に回収ボックスを設けて回答を集めることとしました。

ビジネス目的で訪れている人は、観光目的の人と比べて滞在期間中忙しく、アンケートに回答する人が少ない可能性があります。

この場合、実際よりもビジネス目的の人の割合は低く、観光目的の人の割合は高いという結果が出る可能性があります。



脱落バイアス

脱落バイアス

長期的な調査において、調査の目的や方法に関連する理由でドロップアウトする参加者がいる場合に起こるバイアス。

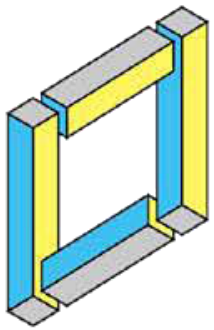
Example

ある慢性疾患の患者に対し、ある治療法の３年後の効果を調査したいとします。

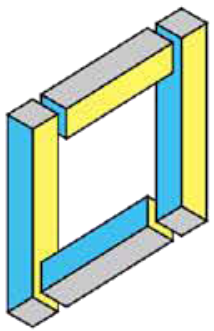
調査開始当初は100人が参加していましたが、そのうち30人が治療の効果を感じられなかったため調査期間の途中で調査協力をやめてしまったとします。

この場合、３年後の調査終了時まで協力し続けていた人を標本として治療法の効果を調べると、効果があった人の割合は

実際よりも高くなってしまう可能性があります。



情報バイアス



情報バイアスの種類

情報バイアスにはたとえば以下のようなものがあります。

- **質問者バイアス**
- **測定バイアス**
- **想起バイアス**
- **検出バイアス**
- **社会的欲求バイアス**

質問者バイアス

質問者バイアス（interviewer bias）

質問者の先入観や態度によって、回答が誘導されたものとなったり、回答の解釈が偏ったりしてしまうことで起こるバイアス。

Example

あるお土産の値段について、値下げを検討している店員が客に対し口頭で質問し調査するとします。

以下の２つの質問の仕方では、前者の方が「高いと思う」という回答が多くなると考えられます。つまり、正確なデータ収集には後者のように中立的な態度で質問を行う必要があります。

「この商品の値段は高いと思いませんか。」

「この商品の値段についてどのように思いますか。」



想起バイアス

想起バイアス（recall bias）

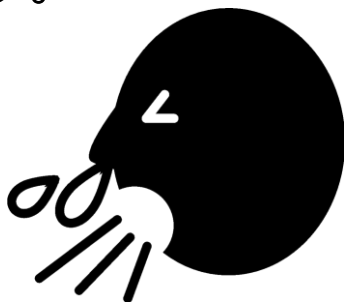
過去の出来事について尋ねたとき、人によって記憶の完全さや正確さが異なることによって引き起こされるバイアス。

Example

子供がくしゃみをしていた頻度について調査しようと、子育てを終えた母親にアンケートを取るとします。

この時、子供が何らかの障害を持っていた母親の方が子供の体調の変化についてより鮮明に覚えていると考えられるため、子供がくしゃみをしていた出来事をよく思い出せる可能性があります。

この場合、「障害のある子供の方がくしゃみをたくさん
していた」というデータが得られてしまう恐れがあります。



検出バイアス

検出バイアス（detection bias）

調査方法の違いにより、調査したいことの発見されやすさが変わってしまって起こるバイアス。

Example

ある地域における1日のコロナウイルスの感染者数を調査したいとします。

PCR検査が毎日ある地域と比べて、PCR検査は希望者に対してのみ任意で実施するという地域では、見つかる感染者数は少なくなると考えられます。

この場合、前者の地域の方が感染者数が多いと結論づけるのは誤りである可能性があります。

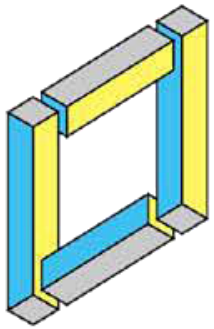


社会的欲求バイアス

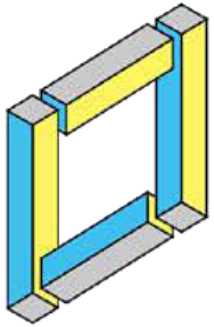
社会的欲求バイアス（social desirability bias）
アンケートなどにおいて自分をより良く見せようとする傾向によって
引き起こされるバイアス。

Example
バイアスの例のスライドで挙げた②の例はまさに社会的欲求バイアス
です。
記名式アンケートであることから、より良い感想を
書こうとしてしまう人が多くなってしまうと考えられます。
このように、記名式などの条件は社会的欲求バイアスを
大きくする原因となるので、調査の方法を設計する際に
意識する必要があります。





まとめ



データ分析以前の着眼点 データの偏り①

・バイアスがあると、せっかく実施した調査結果が信頼できないものとなってしまう。

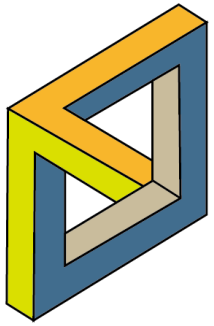
・選択バイアスは母集団から調査対象となる標本を選ぶ際に生じる。

・情報バイアスは測定の仕方に問題がある場合に生じる。

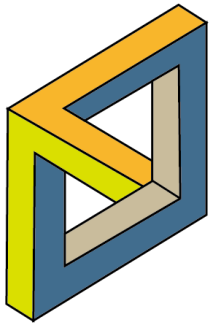
・どのようにバイアスが生じるのかを学ぶことで、実際に調査を行う際にバイアスが発生しにくい方法を選択できるようにすることが重要。

データ分析以前の着眼点

データの偏り ② 偏りの見破り練習



バイアスを探ってみる



バイアスクイズ１

ビール製造メーカーが、スーパーマーケットの出口で、ビールについてのアンケート調査を実施したところ、飲酒ついてネガティブなイメージを持っている回答者が過半数を超えていた。そこで、半年間そのスーパーマーケットで積極的な広報活動を続け、再度アンケート調査をしたところ、ネガティブなイメージを持つ回答者が顕著に減っていた。

この広報活動は成功したといえるだろうか？



バイアスクイズ１


観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

広報活動が成功したかそうでないかを、
個人で考えてみましょう。
その考えを裏付ける理由をなるべくたく
さんあげていきましょう。

調査はどのように行われたかを考えてみましょう。

その調査には、どんなバイアスが潜んでいるのかを

探ってみましょう。

バイアスクイズ１

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

個人で考えたことをグループで共有しましょう。グループで意見を交換して、グループとしての共通見解までたどり着けるようがんばってみましょう。

バイアスクイズ１

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

全体発表で、他のグループの意見を聞き、
自分たちのグループの見解との共通点、
相違点などをとらえていきましょう。

バイアスクイズ２

ある会社で、社内のハラスメント対策についてのアンケート調査を行ったところ、ハラスメントに対しての認知や意識も非常に高いレベルであり、社内でのハラスメント行為はないという回答を得たので、外部組織によるハラスメント対策研修を見送ることにした。

この判断は正しかっただろうか？

バイアスクイズ２


観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

研修を見送ったことが正しかったかどうかを考えてみましょう。
研修を見送った理由として考えられることを、どんな質問に対してどんな回答があったかという観点から考えることで、バイアスを発見できるかもしれません。

調査はどのように行われたかを考えてみましょう。

その調査には、どんなバイアスが潜んでいるのかを探ってみましょう。

バイアスクイズ２

観光ビッグデータ

ワークシート９

バイアスと相関

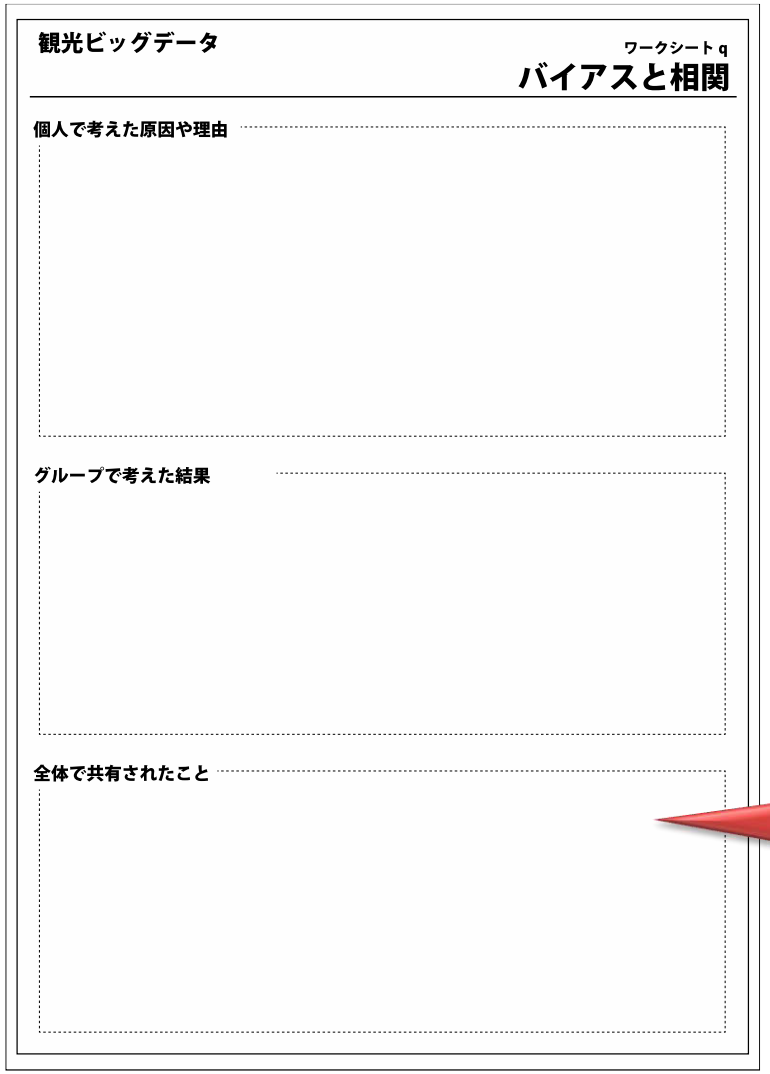
個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

個人で考えたことをグループで共有しましょう。グループで意見を交換して、グループとしての共通見解までたどり着けるようがんばってみましょう。

バイアスクイズ２


観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

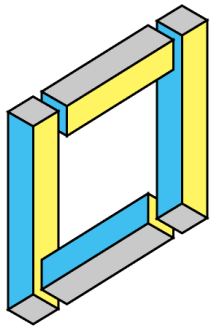
グループで考えた結果

全体で共有されたこと

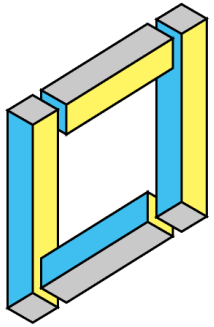
全体発表で、他のグループの意見を聞き、自分たちのグループの見解との共通点、相違点などをとらえていきましょう。

データ分析以前の着眼点

相関と因果①　相関と因果の違い・代表的な混同



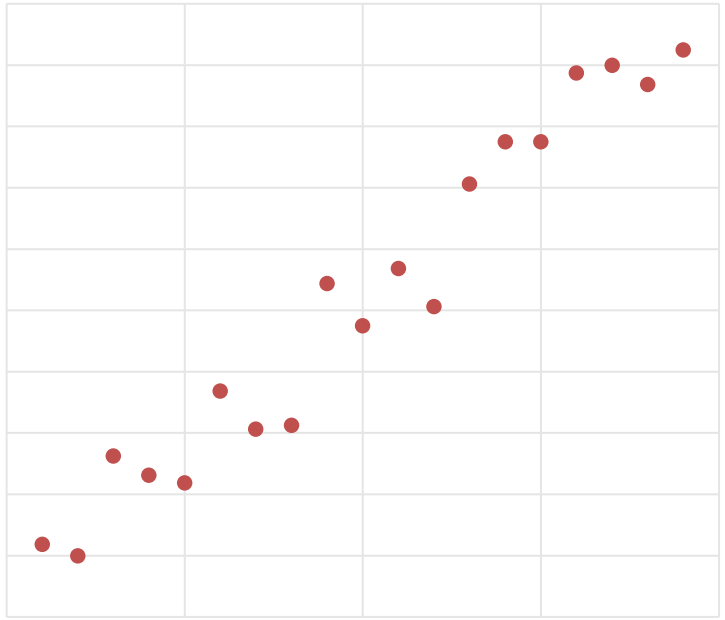
相関とは



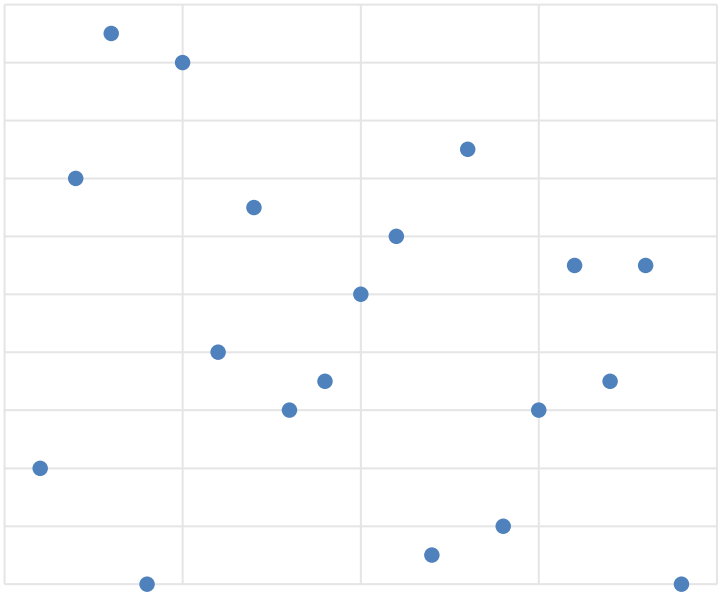
相関とは？

相関：AとBに何らかの関連性がある場合、「AとBに相関がある」と言います。

相関あり



相関なし



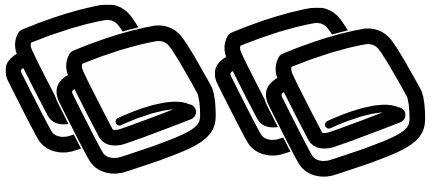
相関の例

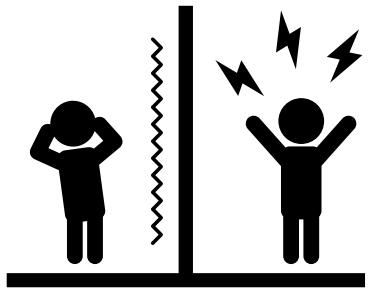
相関には例えば次のようなものがあります。

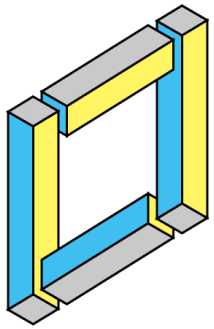
① ある地域で気温が低い日ほど、温泉の利用客が増えると言うデータが得られたとします。
このとき、気温と温泉の利用客数には相関があると言います。



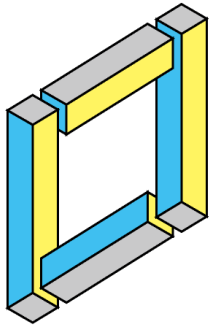
② ある地域でお土産のタオルの売上が多いと、騒音トラブルの件数が増えるというデータが得られたとします。
このとき、タオルの売上と騒音トラブル数には相関があると言います。







因果とは



因果とは？

因果関係

Aを原因としてBという結果になると言える場合、「AとBとの間に因果関係がある」と言います。直接の相関性とも呼ばれます。

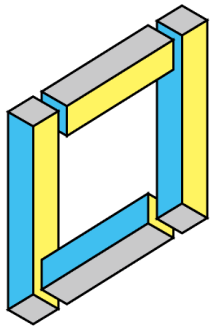
Example

ある山の年間登山者数について、年間降水日数が少ない年ほど登山者が多いと言うデータが得られたとします。

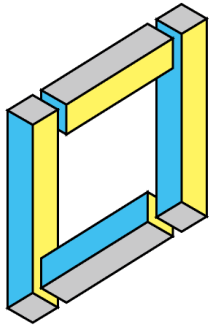
雨天だと登山は危険なので登山客が減ることは容易に想像できます。つまり、降水が原因で登山者数が変化すると言えるでしょう。

このとき、年間登山者数と天気には因果関係があると言えます。





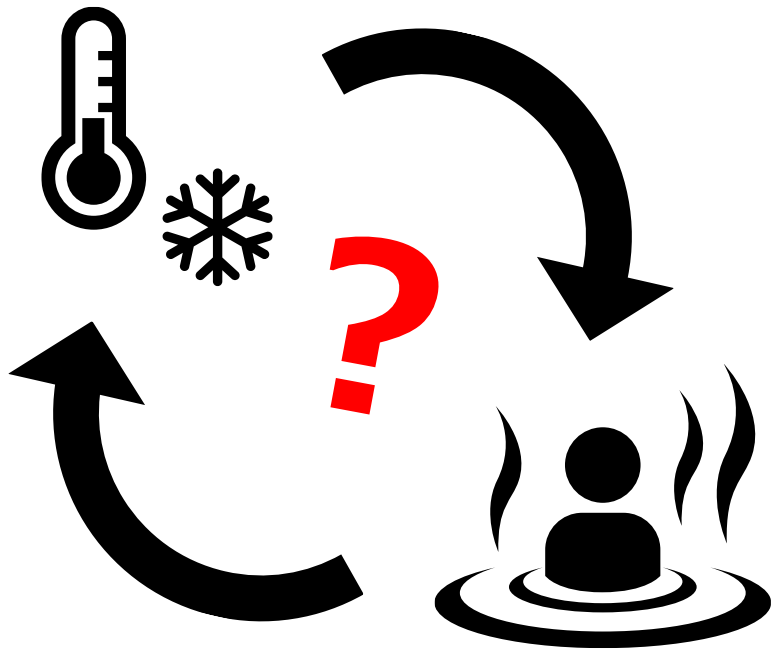
疑似相関とは

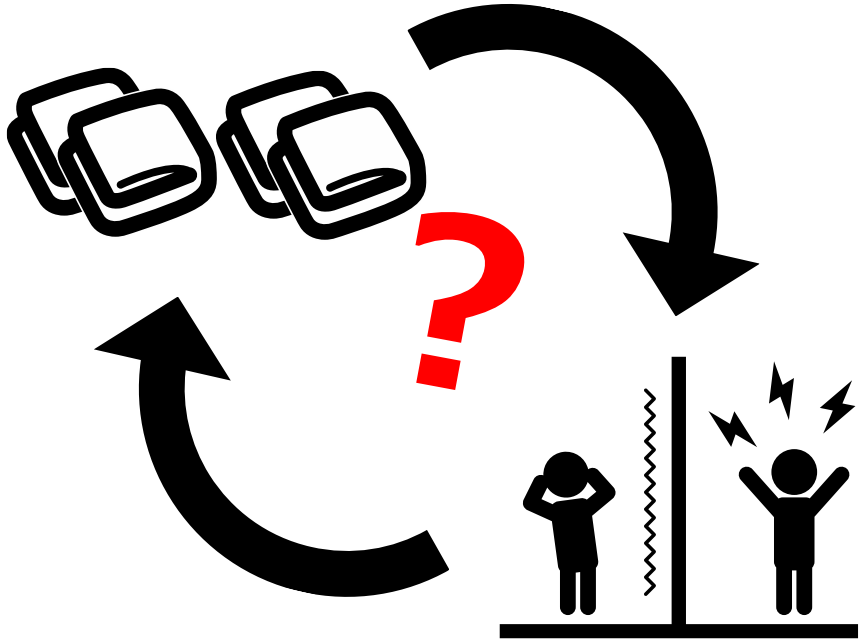


相関と因果

相関の例であげた①と②のそれぞれのケースについて、因果関係があるか考えてみましょう。

ケース①
気温と温泉の利用客数





ケース②
タオルの売上と騒音トラブル数

相関と因果

①に関しては、寒ければ温かい温泉に入る人が増えるというのは妥当なので、因果関係があると言って良いでしょう。

では②はどうでしょうか？

結論から言うと、因果関係はないと考えられます。

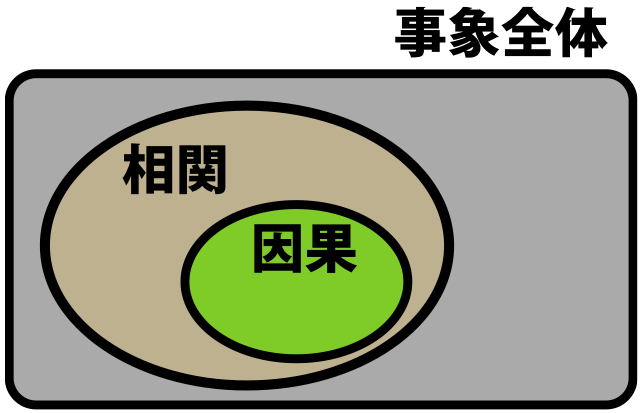
相関はあるため、一見すると、お土産のタオルが売れることで騒音トラブルが増加するように見えます。

しかしよく考えてみると、タオルの売上が騒音トラブル数に影響を与えているとは言い難いでしょう。

このように、

因果があれば相関もありますが、

相関があれば因果があるとは限りません。



疑似相関とは？

疑似相関

２つの事象に直接の相関性（因果関係）がないのに、見えない要因（交絡因子）の存在によってあたかも因果関係があるかのように見える状態のこと。

②はまさに疑似相関の例です。

お土産のタオルの売上が多かったから騒音トラブルが増えたとし、「因果関係がある」とみなしてお土産のタオルの販売を止めるのは誤った判断と言えるでしょう。

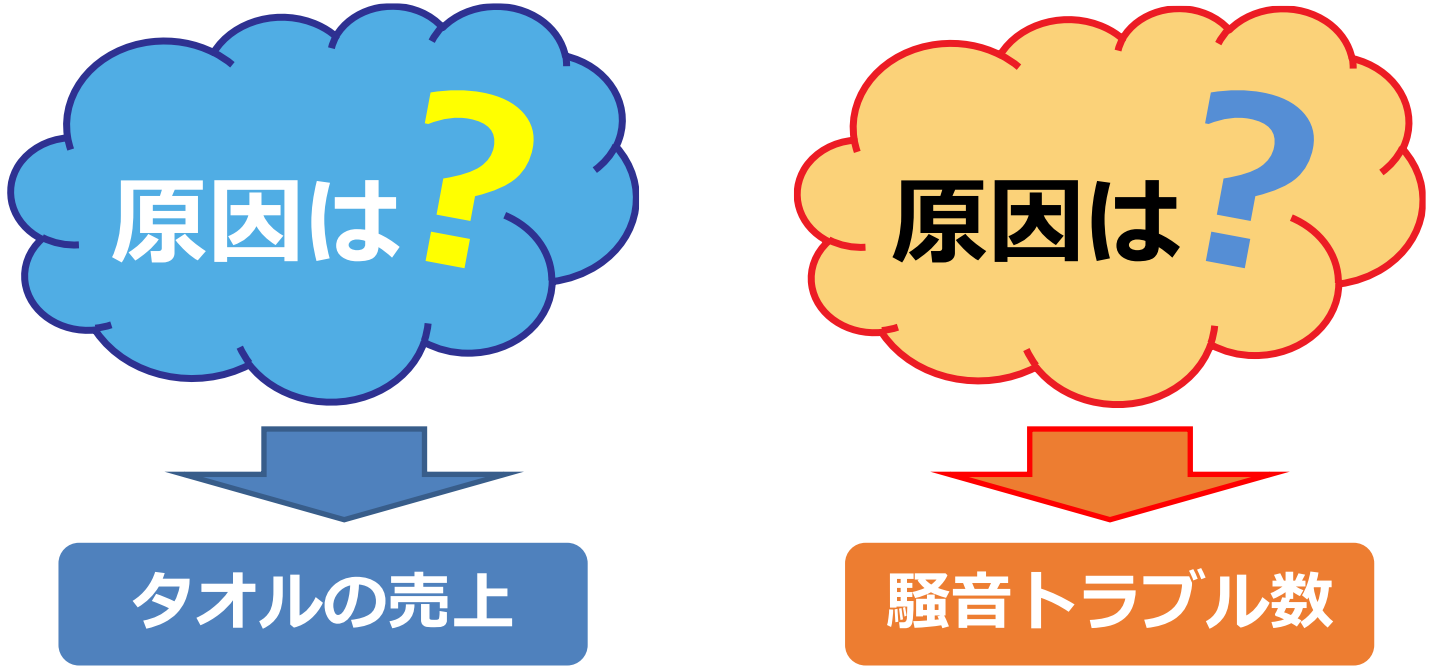
このように「疑似相関がある」場合に「因果関係がある」としてしまわないように注意を払う必要があります。



疑似相関の原因

では、②の例で疑似相関はなぜ生まれたのでしょうか。

お土産のタオルの売上を左右する要因、騒音トラブル数を左右する要因を考えてみましょう。

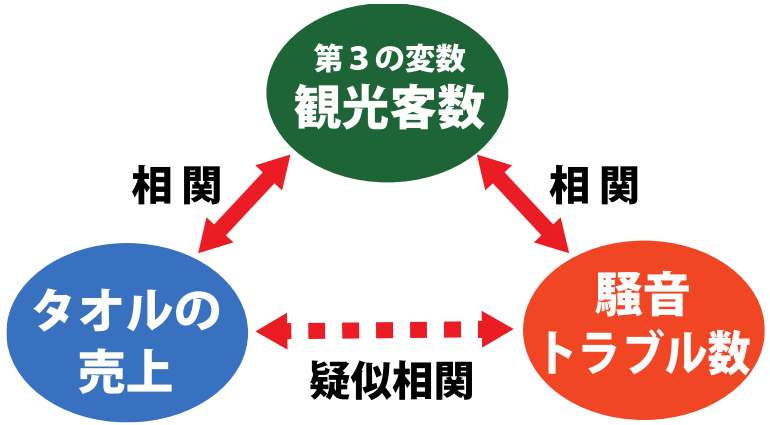


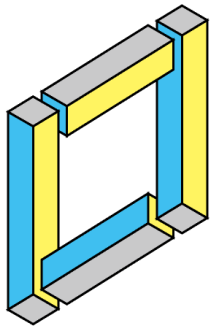
交絡因子とは？

例えば「観光客数」という要素を考えてみるとどうでしょうか。

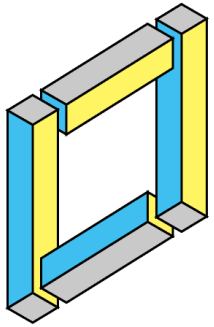
観光客数が増えればお土産のタオルの売上は伸びると考えられます。
また観光客数が増えれば、夜遅くに大声で騒ぐ人が増えたりして、騒音トラブルも増えるでしょう。

この②の例における観光客数のように、
関係を調べたい要素（タオルの売上と
騒音トラブル数）それぞれと相関が
あって、背後で関係性に影響を与える
要素のことを**交絡因子**と言います。





まとめ



データ分析以前の着眼点 相関と因果②

・AとBに何らかの関連性があるとき、AとBには相関がある。

・AがBの結果に影響を与えるとき、AとBには因果関係がある。

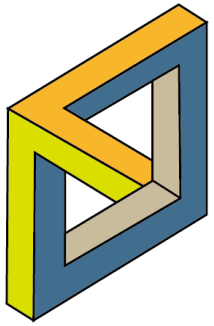
・相関があるからといって因果関係があるとは限らない。

・実際には因果関係がないのにAとBに相関があり、因果関係があるように見えるとき、AとBには疑似相関があると言い、疑似相関を引き起こす要因を交絡因子と言う。

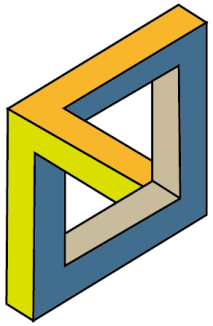
・相関のあるデータを活用するためには、実際に因果関係があるのか、それとも単なる疑似相関なのかを見極め、後者の場合には交絡因子を見つけることが重要。

データ分析以前の着眼点

相関と因果 ② 混同の見破り練習



疑似相関を探ってみる



相関クイズ

ある調査で、血圧の高さと年賀状の枚数に相関があることがわかった。この調査結果を見たある主婦が、夫に年賀状を書くのをやめるようにお願いをした。

夫の血圧は下がっただろうか？



相関クイズ


観光ビッグデータ

ワークシートｑ

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

血圧が下がったかどうかを予想してみましょう。どうしてその予想をたてたかの理由も書き出していきましょう。

血圧の高さと年賀状の枚数には直接の因果があるでしょうか？因果が見つからない場合は、共通する因子を探ってみましょう。

相関クイズ

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

個人で考えたことをグループで共有しましょう。グループで意見を交換して、グループとしての共通見解までたどり着けるようがんばってみましょう。

相関クイズ

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

全体発表で、他のグループの意見を聞き、
自分たちのグループの見解との共通点、
相違点などをとらえていきましょう。

相関ワーク

疑似相関になりそうなものを考えてみましょう。
その際に、かならず共通因子が何であるかについても
一緒に考えてみましょう。

相関ワーク

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

いろいろな擬似相関を考えて書き出してみましょう。
必ずしも相関が証明できるものでなくてもかまいません。

相関ワーク

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

個人で考えたことをグループで共有しましょう。グループの他のメンバーの意見には、個別の議論をせずに、お互いになるべくたくさんの意見が出てくるようにしましょう。

相関ワーク

観光ビッグデータ

ワークシート９

バイアスと相関

個人で考えた原因や理由

グループで考えた結果

全体で共有されたこと

全体発表で、他のグループの発表を聞き、自分たちのグループとの共通点、相違点などをとらえていきましょう。