平均値・中央値・最頻値

作成者：辻　大貴

平均値・中央値・最頻値

- **平均値**

全てのデータの値を足し合わせて、データの数で割ったもの

「42、48、56、62、82」の平均値は、

全てのデータを足し合わせ、データの数で割ると求められる

→全データの合計は「290」、データ数は「5」なので、

「290 ÷ 5」から平均値は「58」になる

平均値・中央値・最頻値

- **中央値**

  データの値を小さい順に並べたときの中央のデータの値のこと

  - **データ数が奇数の場合**：真ん中の値

  - **データ数が偶数の場合**：真ん中に近い2つの値の平均値

平均値・中央値・最頻値

- **中央値（データ数が奇数の場合）**

データが「42、48、56、62、82」の場合、中央値は「56」

42	48	56	62	82

平均値・中央値・最頻値

- **中央値（データ数が偶数の場合）**

データが「42、48、56、62、82、90」の場合、中央値は「59」

42	48	56	62	82	90

平均値・中央値・最頻値

- **最頻値**

最も頻繁に出てくるデータの値のこと

→「1, 1, 1, 1, 1, 1, 4, 4, 5, 8, 10, 13, 15」の場合、最頻値は「1」

値	1	4	5	8	10	13	15
回数	6	2	1	1	1	1	1

平均値・中央値・最頻値

・**外れ値**

他のデータと比べて著しく小さな値、または大きな値のこと

→「42、48、56、62、82、670」の場合、「670」のような値

→平均値は外れ値の影響を受けやすく、

中央値・最頻値は受けにくいという特徴がある

→「42、48、56、62、82、670」の場合、平均値は「160」になる

移動平均

作成者：辻　大貴

移動平均

- **移動平均**

　株価などの時系列のデータを平滑化するときなどに使用される

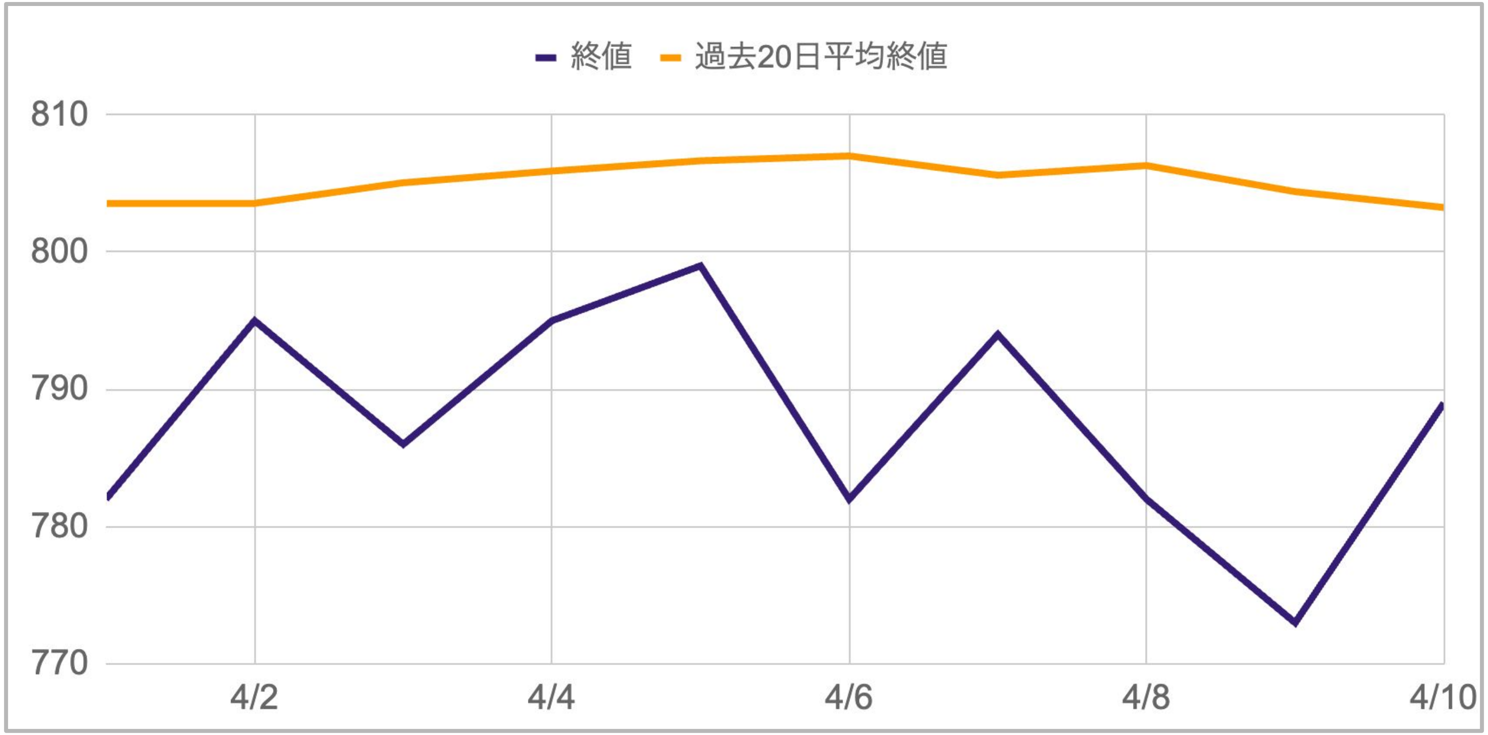
→一定の期間のデータを使用して平均を求めていく手法

→株式の場合、現在の株価以外にも、過去20日の平均株価を

　参考にした方が全体の動きを理解しやすい場合がある

→1日経過したら、新しく20日分のデータを使用して平均を計算

移動平均



移動平均

- **移動平均**

  算術平均（データの値を全て足してデータの数で割ったもの）を使用する方法以外にも**他の平均**を使用する場合がある

→古いデータの値の重みを小さくし、
新しいデータの値の重みを大きくする　など

ヒストグラム

作成者：辻　大貴

ヒストグラム

調査のためにデータを収集したとしても、

データの羅列ではデータ全体の**特徴**を理解することは難しい

- **数学テストの点数**

68、45、90、98、40、73、82、55、60、44

45、83、78、66、63、65、64、35、55、50

59、55、70、67、54、73、82、68、59、58

ヒストグラム

　特徴を理解するために表・グラフで表してみるのが効果的

→グラフにすることで特徴を**可視化**できる

　表やグラフには目的に応じて最適なものを選ぶ

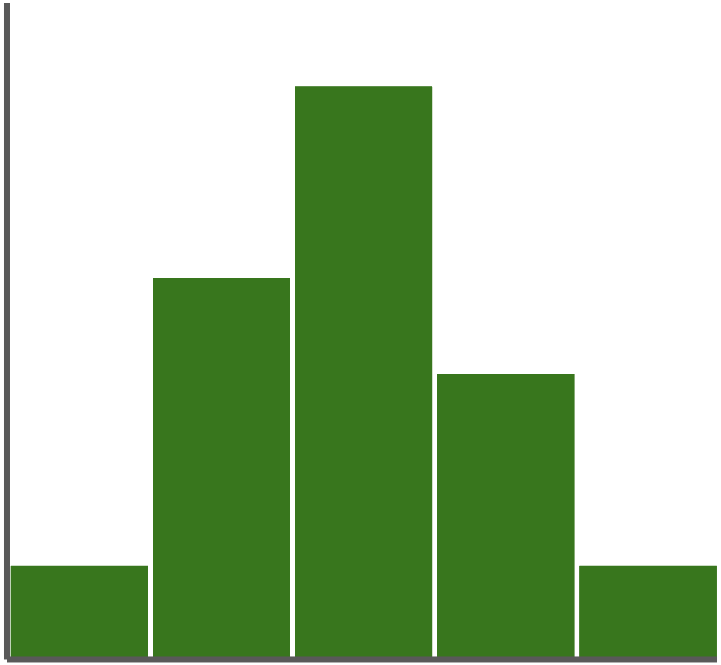
→よく使うグラフもあれば専門家でないと理解できないグラフ

→今回は**ヒストグラム**について解説をしていく

ヒストグラム

- 目標

階級（以上～以下）	度数
３０～４３	２
４４～５７	８
５８～７１	１２
７２～８５	６
８６～９９	２
合計	３０



ヒストグラム

- **ステップ**

データを一定の範囲にグループ分けをしていく

→テストの点数を「30点以上43以下」、「44点以上57以下」

のように、グループに分けていく

→区間の幅はちょうど良い幅を自分で設定していく

→グループ分けしたときのそれぞれの区間を**階級**と呼ぶ

ヒストグラム

・**ステップ**

階級に分けたら、各階級に含まれるデータの個数を数えていく

→「30点以上43以下」に含まれるデータが2個、

「44点以上57以下」に含まれるデータが8個のように

データの個数を1つ1つ数え上げていき、表にまとめていく

ヒストグラム

- ステップ

階級（以上～以下）	度数
３０～４３	2
４４～５７	8
５８～７１	１２
７２～８５	6
８６～９９	2
合計	３０

ヒストグラム

- **ヒストグラム**

横軸に階級、縦軸に度数を取るようにグラフを作成していく

ヒストグラム

- **度数分布**

データを、複数のグループ（階級）に分け、

各グループに属するデータの数（度数）を表したもの

→度数分布を表形式で表現にしたものを**度数分布表**という

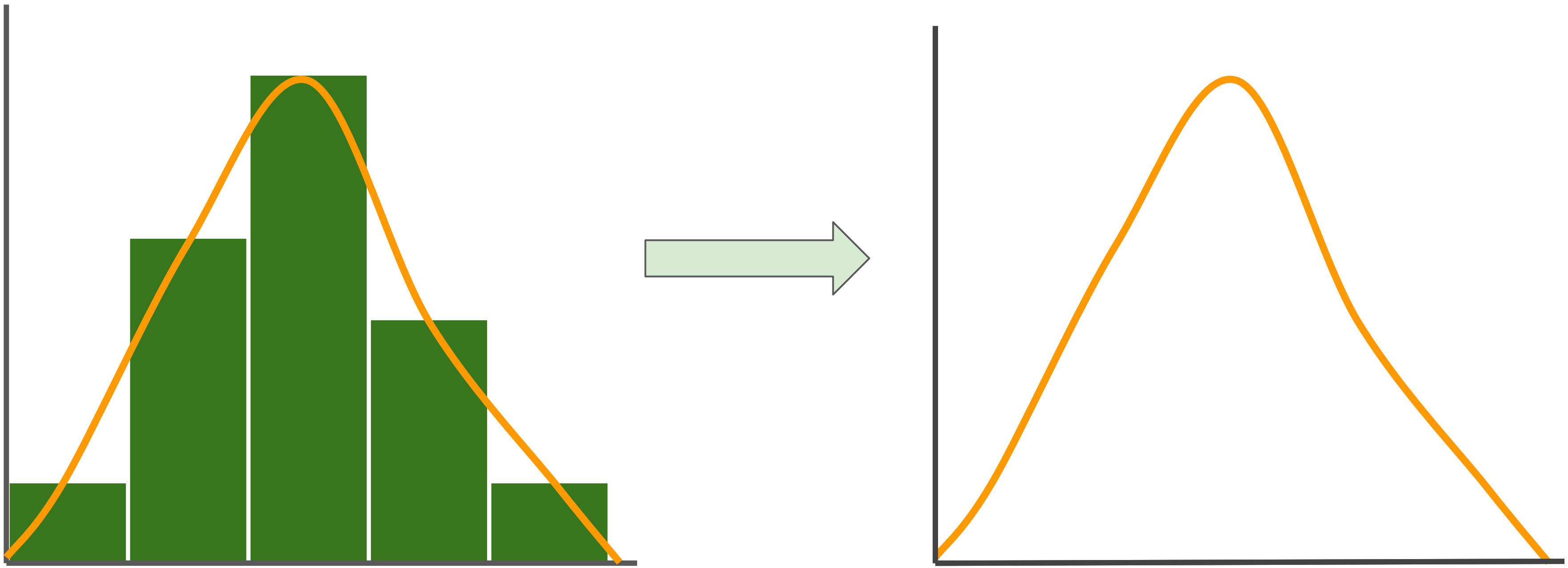
階級（以上～以下）	度数
３０～４３	2
４４～５７	8

ヒストグラムと分布

作成者：辻　大貴

ヒストグラムと分布

- ヒストグラム



ヒストグラムと分布

階級（以上～以下）	度数	確率
３０～４３	2	０.０６７
４４～５７	8	０.２６７
５８～７１	１２	０.４
７２～８５	6	０.２
８６～９９	2	０.０６７
合計	３０	1

ヒストグラムと分布

- **グラフ**

　特定の分布に基づいてランダムにデータを取り出し、

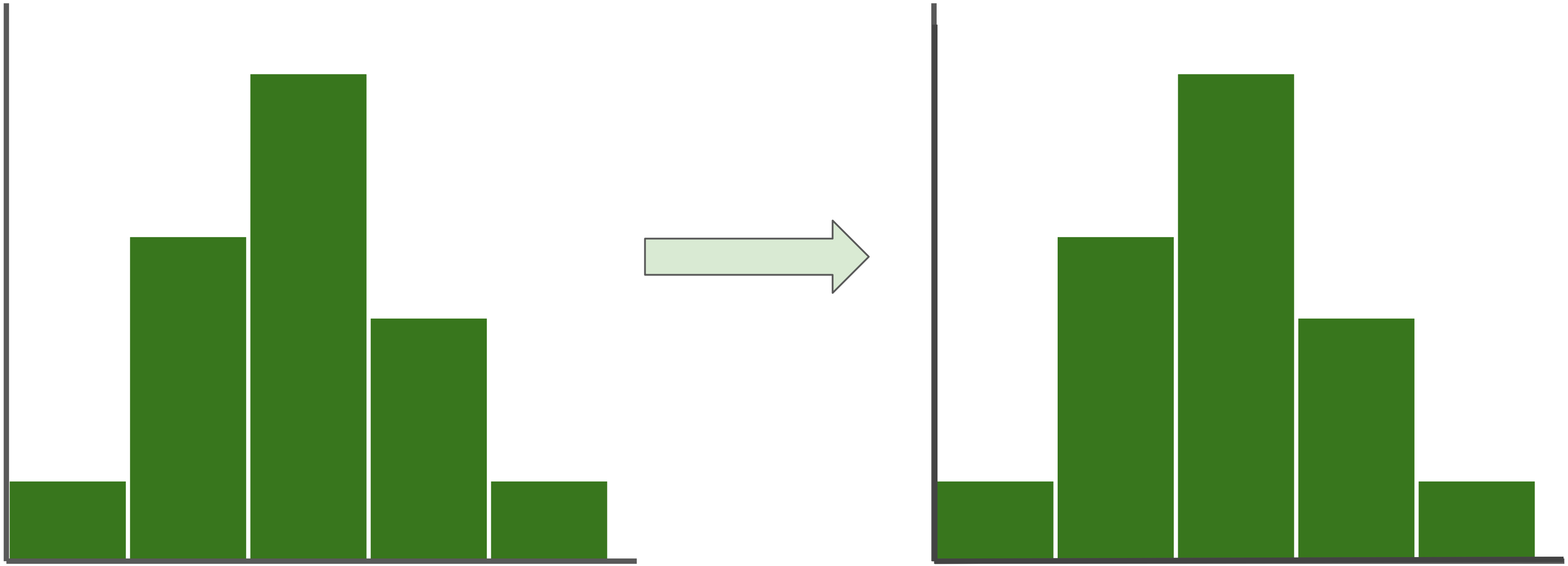
　データを戻して再びランダムにデータを取り出すことを実施

→取り出したデータをもとにしてグラフを作成すると

　縦軸の度数は異なるが同じ形の分布になるという特徴がある

ヒストグラムと分布

- ヒストグラム



分散・標準偏差

作成者：辻　大貴

分散・標準偏差

- **分散**

　各データの偏差を2乗した和をデータ数で割ったもの

→各データの偏差の2乗の平均値が分散である

- **偏差**

　データから平均値を引いた値のこと

→データAが50、平均値が30ならば「50 - 30」から偏差は20

分散・標準偏差

氏名	点数	偏差	偏差の２乗
Aさん	４０		
Bさん	６０		
Cさん	８０		
Dさん	３０		
Eさん	４０		
平均		合計	

分散・標準偏差

氏名	点数	偏差	偏差の２乗
Aさん	４０		
Bさん	６０		
Cさん	８０		
Dさん	３０		
Eさん	４０		
平均	５０	合計	

分散・標準偏差

氏名	点数	偏差	偏差の２乗
Aさん	４０	ー１０	
Bさん	６０	１０	
Cさん	８０	３０	
Dさん	３０	-２０	
Eさん	４０	-１０	
平均	５０	合計	

分散・標準偏差

氏名	点数	偏差	偏差の２乗
Aさん	４０	－１０	１００
Bさん	６０	１０	１００
Cさん	８０	３０	９００
Dさん	３０	-２０	４００
Eさん	４０	-１０	１００
平均	５０	合計	

分散・標準偏差

氏名	点数	偏差	偏差の２乗
Aさん	４０	ー１０	１００
Bさん	６０	１０	１００
Cさん	８０	３０	９００
Dさん	３０	-２０	４００
Eさん	４０	-１０	１００
平均	５０	合計	１６００

分散・標準偏差

- **分散**

各データの偏差を2乗した和が1600、データ数が5なので

「$1600 \div 5$」から「320」が分散になる

- **標準偏差**

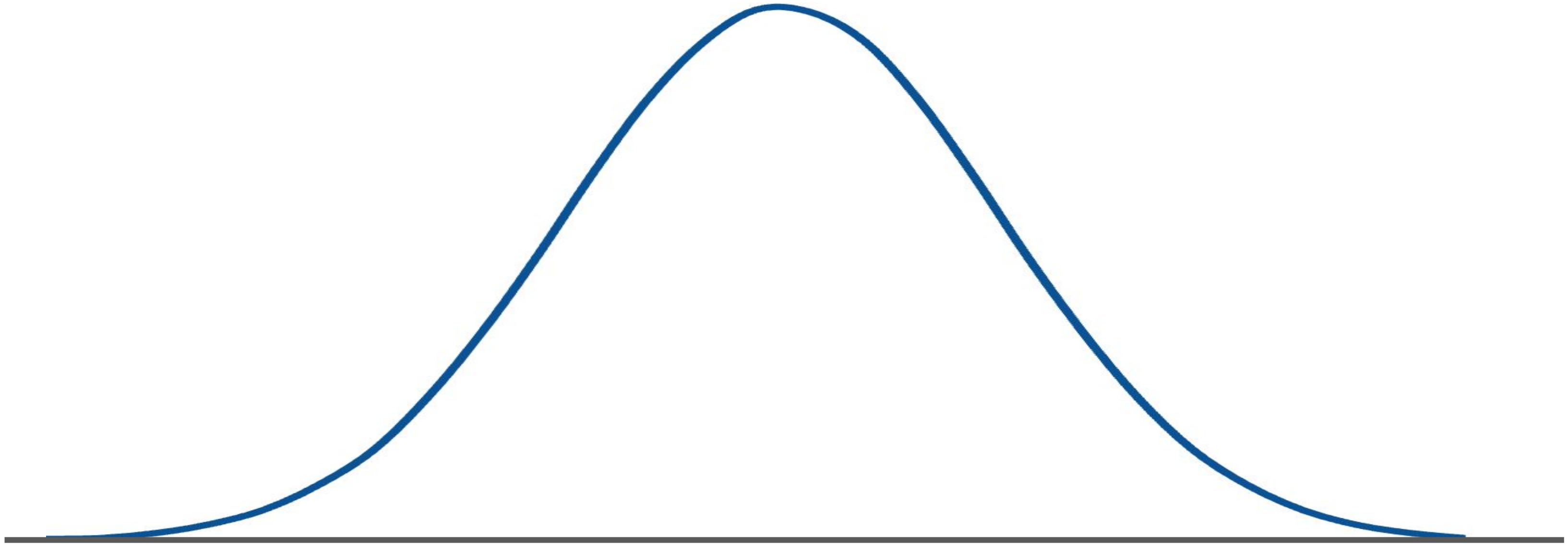
分散の平方根のことで、$\sqrt{320}$ から約17.89が標準偏差になる

正規分布と標準偏差

作成者：辻　大貴

正規分布と標準偏差

正規分布とは左右対称の分布のこと



正規分布と標準偏差

- **正規分布の例**

身長、体重、学力、製品のサイズ誤差　など

→データの平均をμ、標準偏差をσと表すと以下の関係が成り立つ

$\mu - \sigma$から $\mu + \sigma$の間に入る確率は約68%

$\mu - 2\sigma$から$\mu + 2\sigma$の間に入る確率は約95%

$\mu - 3\sigma$から$\mu + 3\sigma$の間に入る確率は約99.7%

正規分布と標準偏差

- **正規分布の例**

テストの平均が50点、標準偏差が17.89の場合

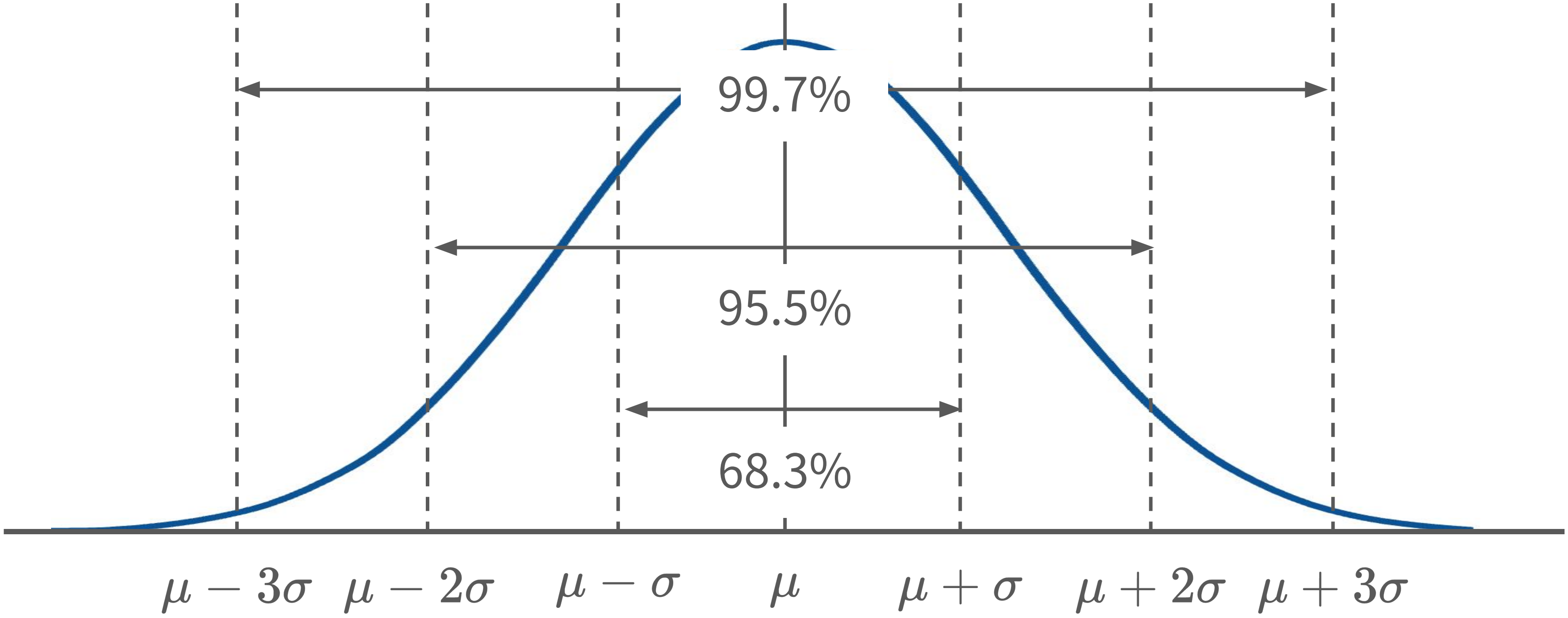
→ $\mu - \sigma$から $\mu + \sigma$の間に入る確率は約68%になる

→32.11～67.89点に約68%の人が存在する計算になる

正規分布と標準偏差

平均　　：μ

標準偏差：σ



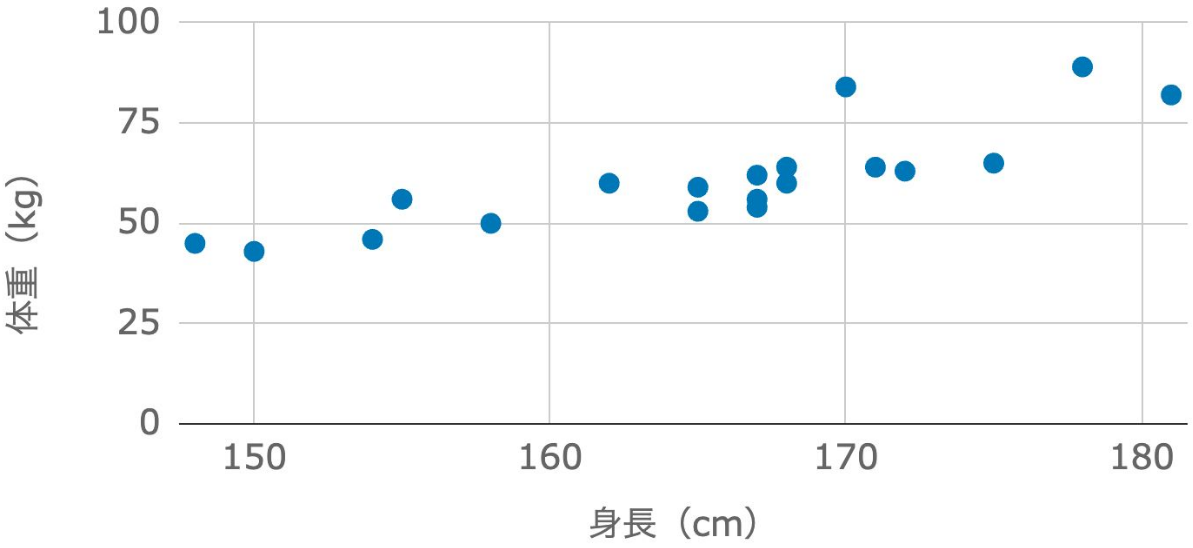
相関関係と相関係数

作成者：辻　大貴

相関関係と相関係数

- **相関関係**

  データ間に存在する関連性のこと



相関関係と相関係数

- **正の相関**

　１つのデータの値が増加すれば、もう一つのデータの値も増加

- **負の相関**

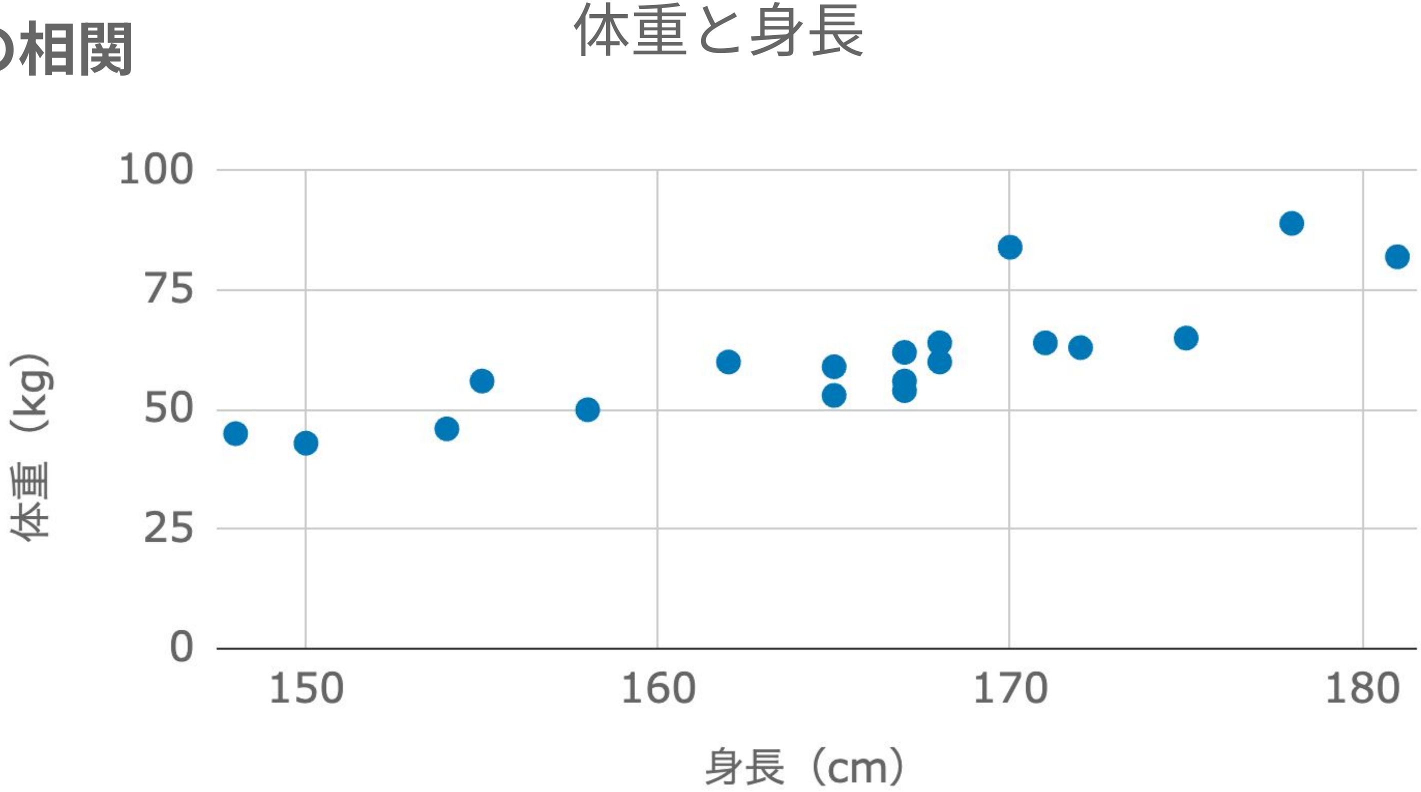
　１つのデータの値が増加すれば、もう一つのデータの値は減少

- **無相関**

正の相関も負の相関もないもの

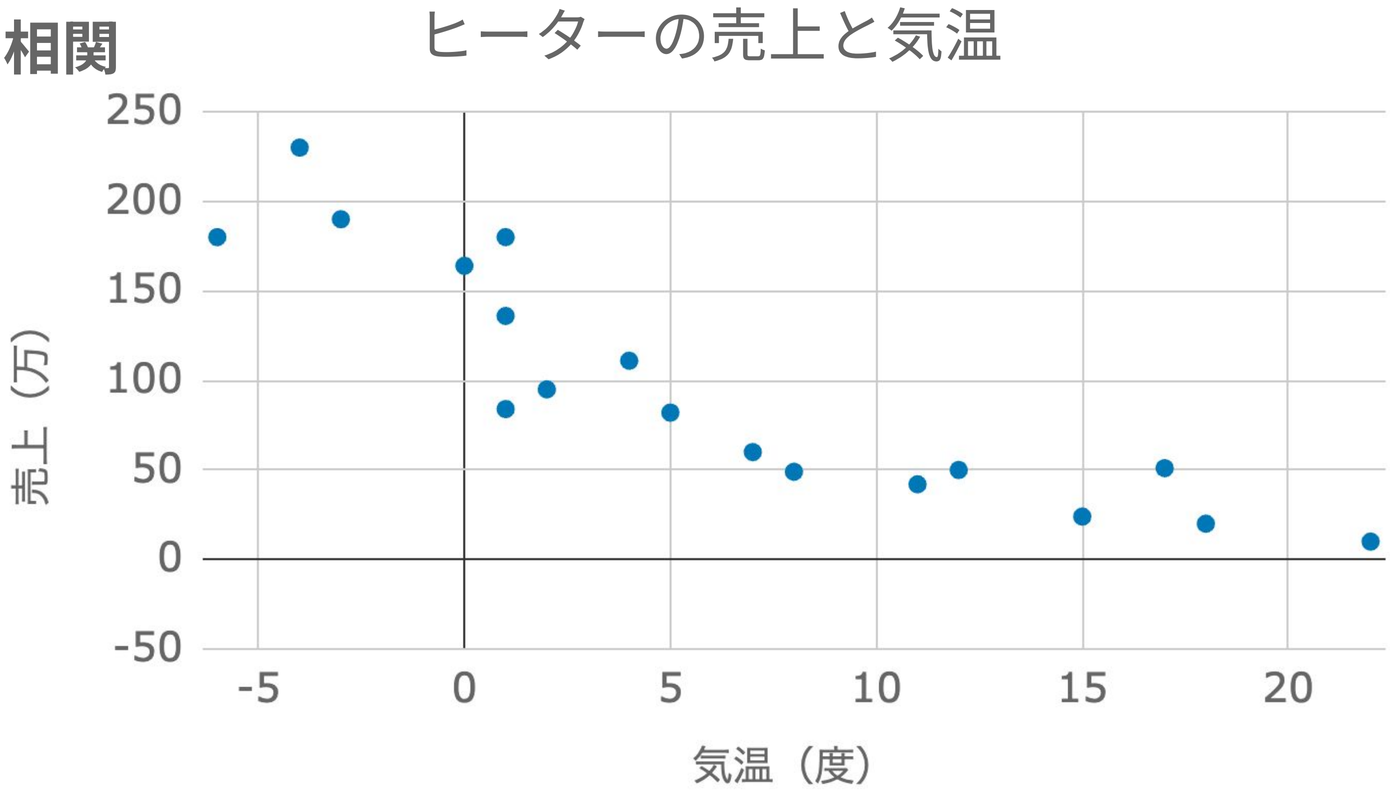
相関関係と相関係数

- **正の相関**



相関関係と相関係数

- 負の相関



相関関係と相関係数

- **強い相関**

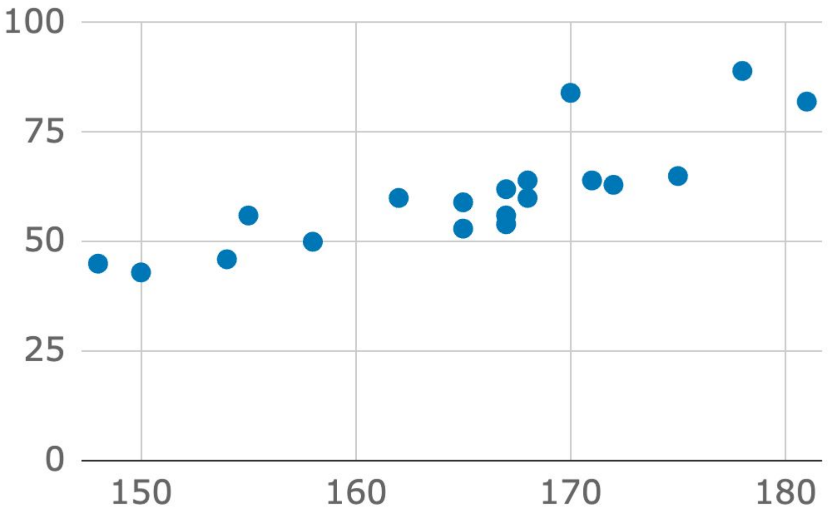
  データ間の関連性が強い相関のこと

- **弱い相関**

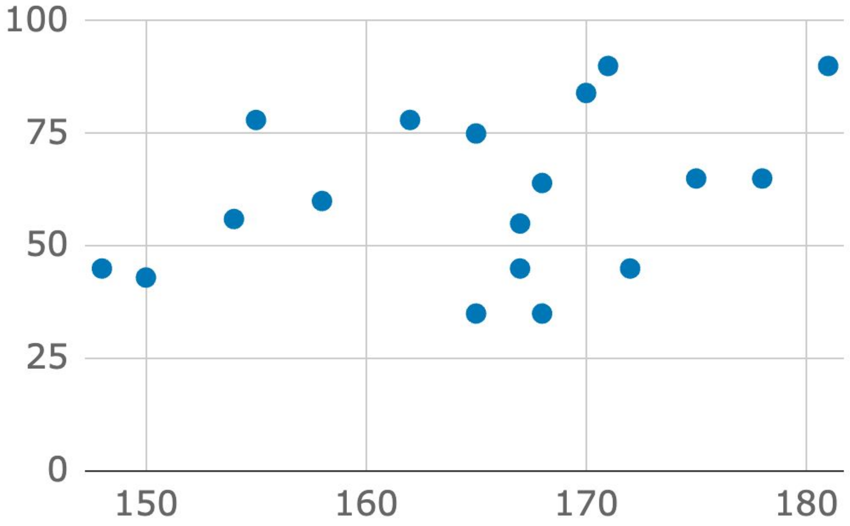
  データ間の関連性が弱い相関のこと

相関関係と相関係数

強い相関



弱い相関



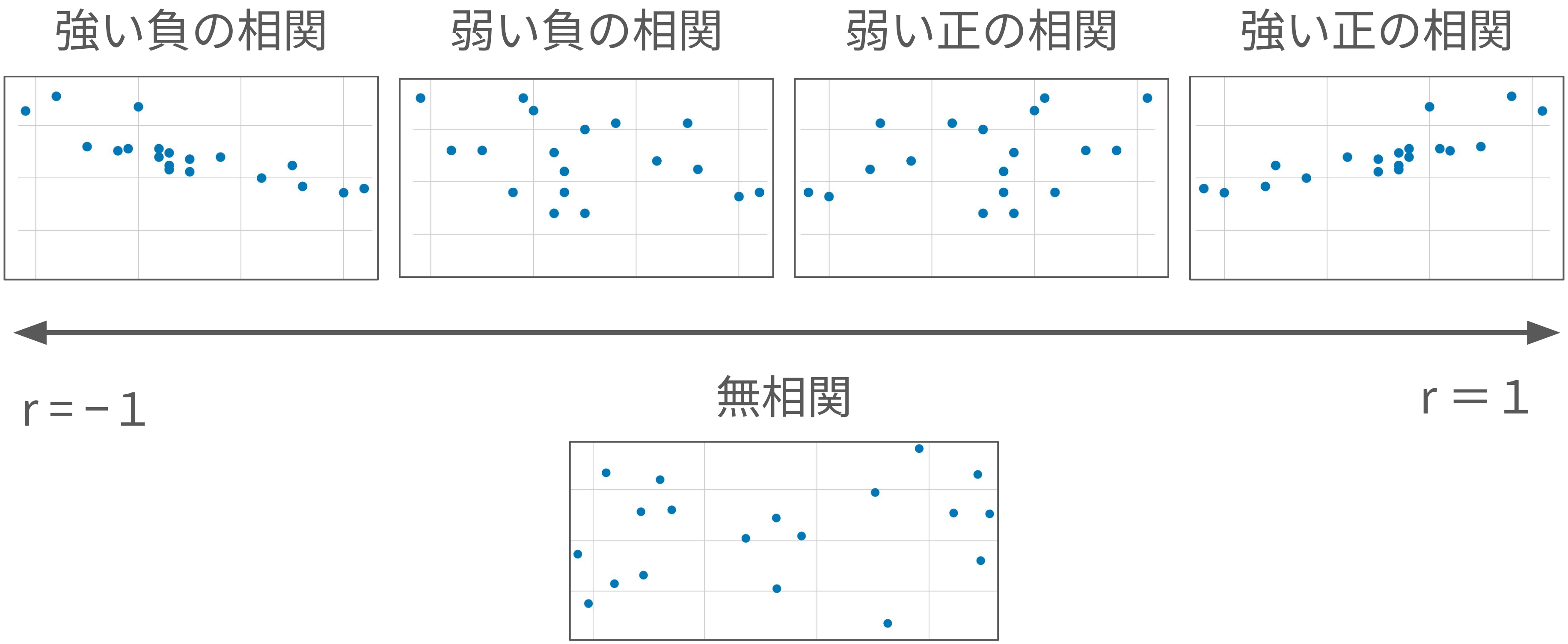
相関関係と相関係数

- **相関係数**

  データの相関関係の強さを表す値のこと

  - 相関係数は**-1～1**の値をとる
  - 相関係数が1に近ければ正の相関
  - 相関係数が–1に近ければ負の相関
  - 相関係数が0に近いほど関係が弱い
  - 相関係数が-1または1に近いほど関係が強い

相関関係と相関係数



相関係数の計算

作成者：辻　大貴

相関係数の計算

- **相関係数**

  データの相関関係の強さを表す値のこと

$$相関係数(r) = \frac{x, yの共分散}{xの標準偏差 \times yの標準偏差}$$

- **共分散**

  2つのデータの偏差の積の平均値のこと

相関係数の計算

	身長(x)	体重(y)	偏差（身長）	偏差（体重）	偏差積
1	172	62	0	-2	0
2	162	58	-10	-6	60
3	180	82	8	18	144
4	178	58	6	-6	-36
5	168	60	-4	-4	16
平均	172	64		合計	184

相関係数の計算

- **共分散**

$184 \div 5 = 36.8$

- **相関係数**

$$相関係数(r) = \frac{x, yの共分散}{xの標準偏差 \times yの標準偏差}$$

相関係数の計算

	身長 (x)	偏差 (身長)	偏差の２乗 (身長)
1	172	0	0
2	162	-10	100
3	180	8	64
4	178	6	36
5	168	-4	16
平均	172	分散	43.2

相関係数の計算

	体重 (y)	偏差 (体重)	偏差の２乗 (体重)
1	62	-2	4
2	58	-6	36
3	82	18	324
4	58	-6	36
5	60	-4	16
平均	64	分散	83.2

相関係数の計算

- **分散**

身長の分散：43.2　→　標準偏差：6.6

体重の分散：83.2　→　標準偏差：9.1

共分散　　　：36.8

$$\text{相関係数}(r) = \frac{x, y\text{の共分散}}{x\text{の標準偏差} \times y\text{の標準偏差}} = \frac{36.8}{6.6 \times 9.1} \fallingdotseq 0.61$$

擬似相関

作成者：辻　大貴

擬似相関

- **因果関係**

  原因と結果の関係を意味する用語

→**相関関係**はデータ間に存在する関連性のこと

→相関関係があったとしても因果関係があるわけではない

→相関関係のように見えるが因果関係ではない関係のことを

擬似相関（見かけ上の相関）と呼ばれる

擬似相関

　ビールの売上とアイスの売上は擬似相関の有名な例

→気温が上がるとビールの売上は上がる

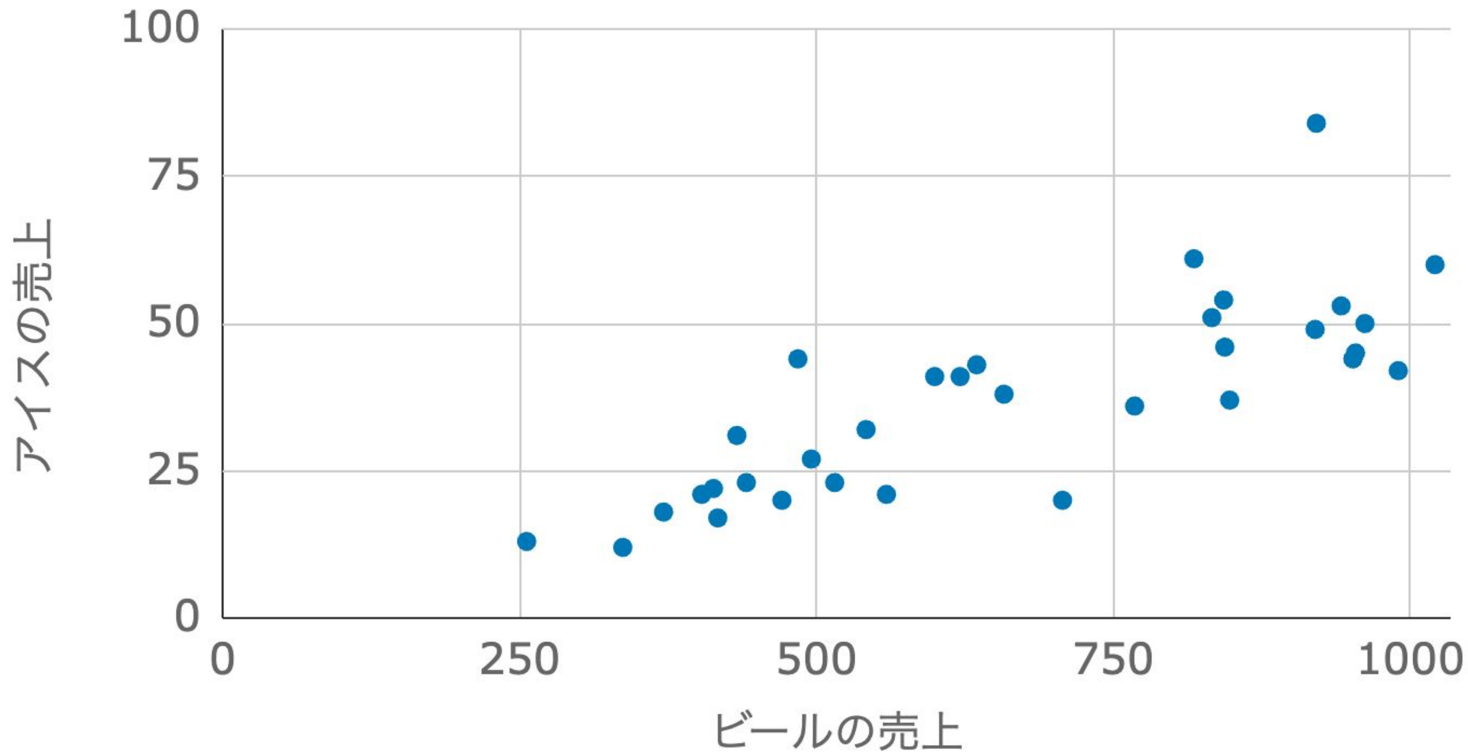
　気温が上がるとアイスの売上は上がる

→見方によってはビールの売上が上がると

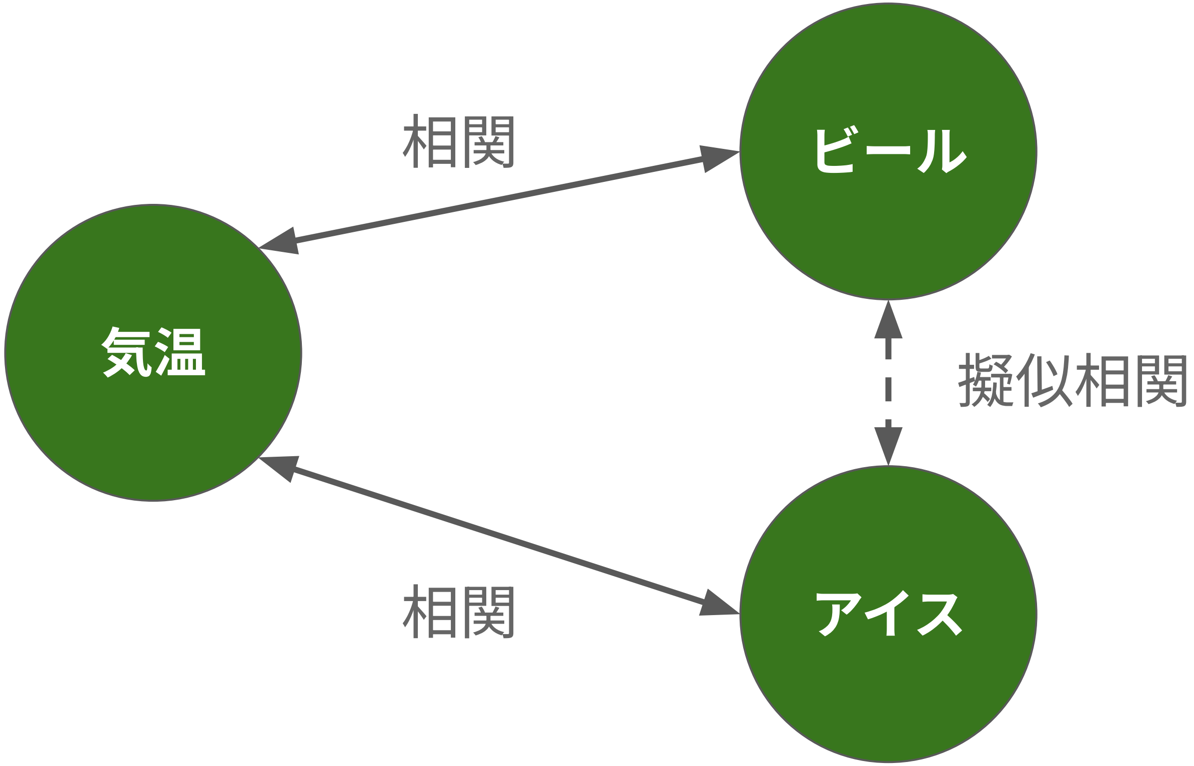
　アイスの売上が上がっているように見える

→因果関係はないが、相関関係は存在してしまう

擬似相関



擬似相関



偏相関係数

作成者：辻　大貴

偏相関係数

- **偏相関係数**

  特定の変数の影響を取り除いた2変数の相関係数のこと

→相関係数と同様に-1〜1の値をとる

→1に近いほど正の相関、-1に近いほど負の相関、
0に近いほど相関なしである

偏相関係数

- **偏相関係数**

変数x、y、zがあり、zの影響を取り除いた変数x、yの偏相関係数

$$偏相関係数 = \frac{\text{xとyの相関係数} - \text{xとzの相関係数} \times \text{yとzの相関係数}}{\sqrt{1 - \text{xとzの相関係数}^2} \times \sqrt{1 - \text{yとzの相関係数}^2}}$$

→各変数の相関係数が分かれば偏相関係数を求めることができる

偏相関係数

- **偏相関係数**

  ビールの売上（x）、アイスの売上（y）、気温（z）のとき、

  zの影響を取り除いた変数x、yの偏相関係数を求めることで、

  疑似相関かどうかを判断することができる

→偏相関係数が0に近い値なら疑似相関である可能性が高い

偏相関係数が-1や1に近い値なら疑似相関でない可能性が高い

偏相関係数

- **偏相関係数**

xとyの相関係数が0.82、xとzの相関係数が0.85

yとzの相関係数が0.88のとき、zの影響を除いたx、yの偏相関係数

$$\text{偏相関係数} = \frac{\text{xとyの相関係数} - \text{xとzの相関係数} \times \text{yとzの相関係数}}{\sqrt{1 - \text{xとzの相関係数}^2} \times \sqrt{1 - \text{yとzの相関係数}^2}}$$

$$= \frac{0.82 - 0.85 \times 0.88}{\sqrt{1 - 0.85^2} \times \sqrt{1 - 0.88^2}} \fallingdotseq 0.29$$

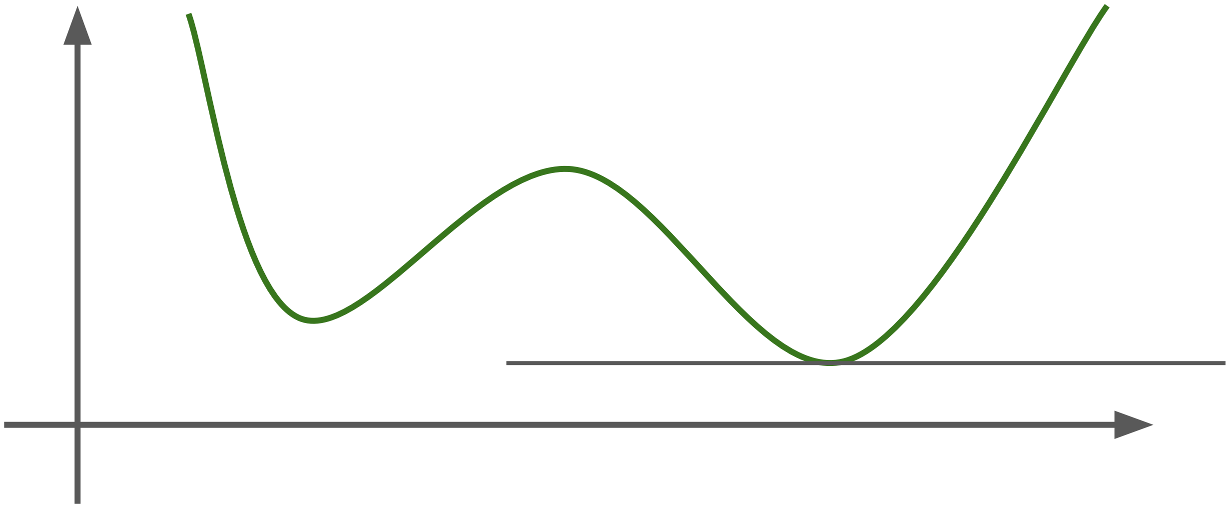
偏微分

作成者：辻　大貴

偏微分

- **微分**

関数の変化の割合（傾き）を求めること



偏微分

- 微分の公式

$$(x^a)' = ax^{a-1}$$

$$(x^3)' = 3x^{3-1}$$
$$= 3x^2$$

$$(nx^a)' = anx^{a-1}$$

$$(2x^4)' = 2 \times 4x^{4-1}$$
$$= 8x^3$$

$$y = 2x^4$$
$$(y)' = (2x^4)'$$
$$= 2 \times 4x^{4-1}$$
$$= 8x^3$$

偏微分

- 微分の公式

$$(x^a)' = ax^{a-1}$$

$$\begin{aligned}(x^3)' &= 3x^{3-1}\\ &= 3x^2\end{aligned}$$

$$\begin{aligned}(2x^4)' &= 2 \times 4x^{4-1}\\ &= 8x^3\end{aligned}$$

$$y = 2x^4 + x^3 + 6$$

$$\begin{aligned}(y)' &= \left(2x^4 + x^3 + 6\right)'\\ &= 2 \times 4x^{4-1} + 3x^{3-1}\\ &= 8x^3 + 3x^2\end{aligned}$$

定数項を微分すると0になる

偏微分

- **偏微分**

1つの変数に注目して微分をすることが**偏微分**という

→$y = 3a^3 + 8a^2b^3 - b^2c^2$のとき、$a$で微分をすると

$$\frac{\Delta y}{\Delta a} = 3 \times 3a^{3-1} + 8 \times 2a^{2-1}b^3$$

$$= 9a^2 + 16ab^3$$

→残りの変数は定数として考える（ $y = 3a^3 + 8a^2\boxed{b^3} - \boxed{b^2c^2}$）

偏微分

- **偏微分**

1つの変数に注目して微分をすることが**偏微分**という

→$y = 3a^3 + 8a^2b^3 - b^2c^2$のとき、$b$で微分をすると

$$\frac{\Delta y}{\Delta b} = 8 \times 3a^2b^{3-1} - 2b^{2-1}c^2$$

$$= 24a^2b^2 - 2bc^2$$

→残りの変数は定数として考える（ $y = \boxed{3a^3} + 8\boxed{a^2}b^3 - b^2\boxed{c^2}$）