度数分布表

データの分布を
見やすくした！

食べログの評価	人数
0 − 1	5
1 − 2	7
2 − 3	8
3 − 4	4
4 − 5	1

以上　未満

→ 変量
→ 度数

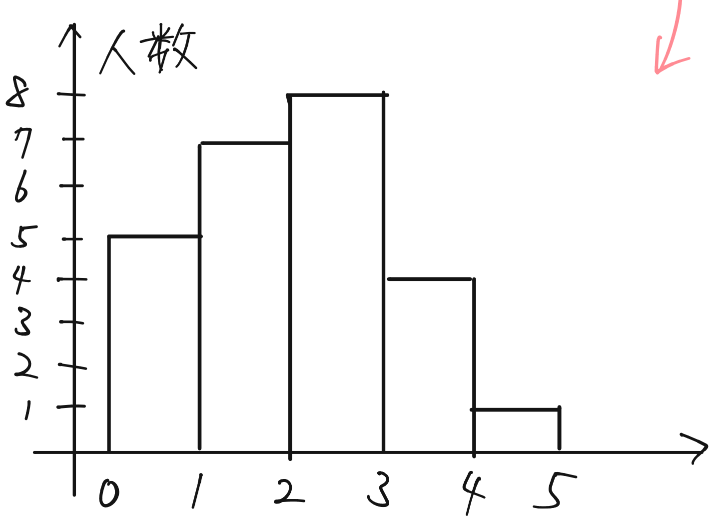
階級

階級の幅：1
階級値：0.5，1.5，…
真ん中の値

ヒストグラム

さらに見やすく
視覚的に．

表から
グラフへ



相対度数

評価（変量）	0−1	1−2	2−3	3−4	4−5
人数（度数）	5	7	8	4	1
相対度数	0.20	0.28	0.32	0.16	0.04
累積相対度数	0.20	0.48	0.80	0.96	1.00

全体
25人

割合

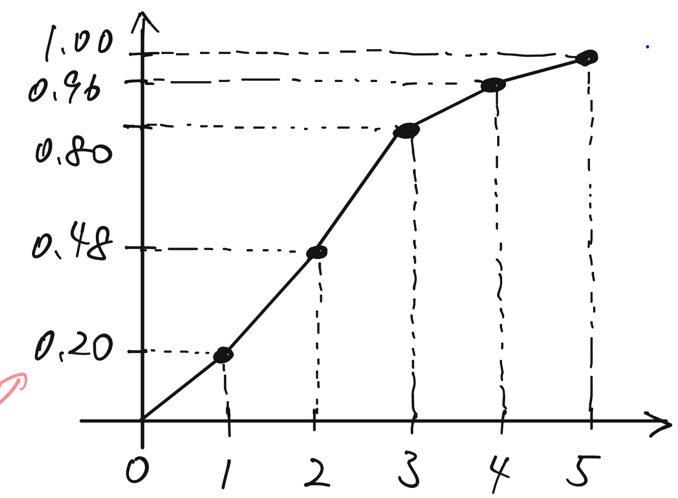
相対度数を足していく．

累積相対度数（全体の中の割合を積み上げたもの）
をグラフで見やすくしたもの

→ 累積相対度数折れ線

折れ線で結んだだけ．

累積相対度数



代表値

平均値 → $\dfrac{\text{データの総和}}{\text{データの個数}}$

中央値 → 小さい順にデータを並べたときの真ん中の値
(偶数個のときは2つの平均値)

最頻値 → データの中で最もよく出てくる値
(度数分布表では、度数が最も多い階級の階級値)

第1四分位数 → 小さい順に並べたときに25%目くらいの値

第3四分位数 → 小さい順に並べたときに75%目くらいの値

<教科書の定義>

① データの中央値を境に、データを2つのデータに分ける。
(2つの個数は等しくなる)

偶数個のとき
$\underbrace{○○\cdots○}\,|\,\underbrace{○\cdots○○}$

奇数個のとき。
$\underbrace{○○\cdots○}\,\boxed{○}\,\underbrace{○\cdots○○}$

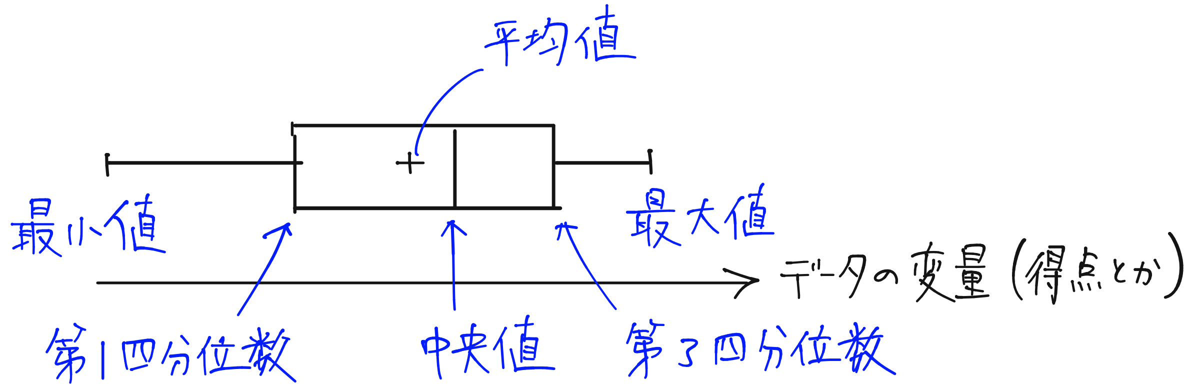
② 左のデータの中央値が第1四分位数

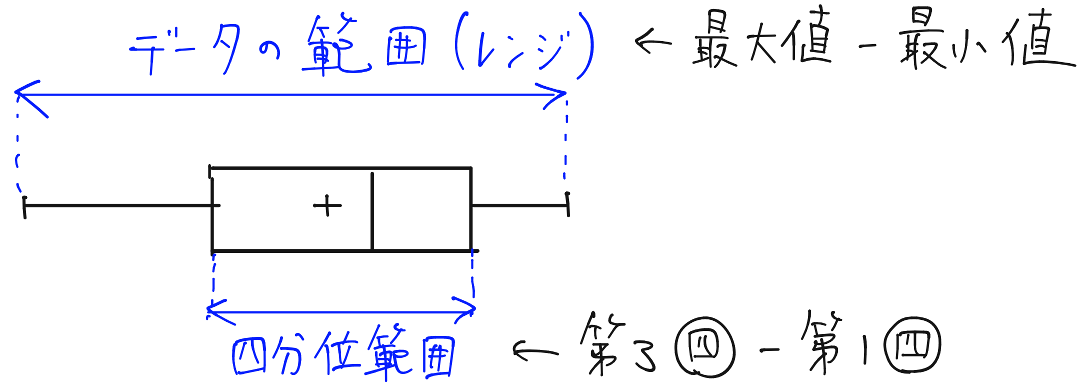
③ 右のデータの中央値が第3四分位数

箱ひげ図

データの分布を見やすくした!

ヒストグラムより他のデータとの比較がしやすい。(並べやすい)





四分位偏差 = $\dfrac{\text{四分位範囲}}{2}$

基準からのズレ

分散と標準偏差

平均からどれくらいデータが散らかっているかを表す.

データを $x_1, x_2, \cdots, x_n$ とおく。(← たとえば得点)

平均値 $\overline{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ ← データの総和 ← データの個数

(平均値からの) 偏差 $= x_1 - \overline{x},\ x_2 - \overline{x},\ \cdots,\ x_n - \overline{x}$

つまり, ズレ

2乗の平均

分散 $s^2 = \dfrac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + \cdots + (x_n-\overline{x})^2}{n}$

標準偏差 $s = \sqrt{s^2}$ ← 分散の正の平方根

$$= \sqrt{\dfrac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + \cdots + (x_n-\overline{x})^2}{n}}$$

☆単位が元データとそろってうれしい. 標準化 $z = \dfrac{x-\overline{x}}{s}$ とか

※度数分布から平均値・分散を求めるときは,
度数の数だけ階級値のデータがあると
思えばOK.

階級値	x_1	x_2	$\cdots$	x_n
度数	f_1	f_2	$\cdots$	f_n

x_1 が f_1 コ　　x_n が f_n コ

平均値 $\overline{x} = \dfrac{x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots + x_n \cdot f_n}{n}$

分散 $s^2 = \dfrac{(x_1-\overline{x})^2 \times f_1 + \cdots + (x_n-\overline{x})^2 \times f_n}{n}$

個数分かける!

☆ 分散の別の求め方

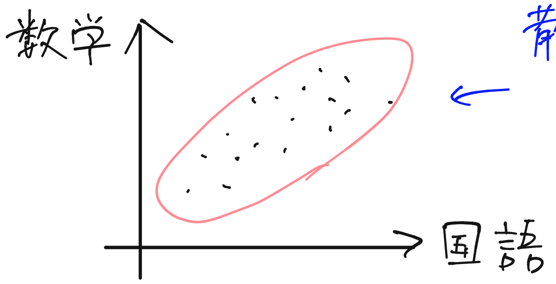
x の分散 $=$ x^2 の平均値 $-$ (x の平均値)^2

ときどきめっちゃラクになる.

相関

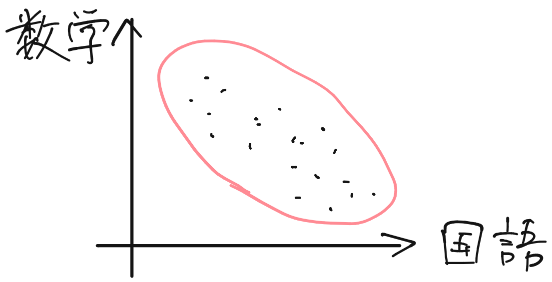
2種類のデータの間の関係。
国語と数学のクラスの得点分布とか。

正の相関関係



散布図 ← 点は1人1人を表す。

負の相関関係



↓

なんか値で測りたい。

データを (x_1, y_1) (x_2, y_2) … (x_n, y_n) とおき。

A くんの国語　数学とか。

x, y のデータの平均値を $\bar{x}, \bar{y}$ とする。

共分散 $= \dfrac{(x_1-\bar{x})(y_1-\bar{y}) + \cdots + (x_n-\bar{x})(y_n-\bar{y})}{n}$

2つのデータをくみ合わせた分散のようなもの。

これが ⊕ だと正の相関
⊖ だと負の相関

↓

値の大小比較がまだできない。

相関係数 $= \dfrac{x\text{と}y\text{の共分散}}{x\text{の標準偏差} \times y\text{の標準偏差}}$

→ −1〜1の、単位のない値となる！

相関の大きさの比較ができる！

変量の変換

2つの変量（得点とか）x, y の間に $y = ax + b$ という関係が成り立つとき、

yの平均値 $= a \times$ (xの平均値) $+ b$

yの分散 $= a^2 \times$ (xの分散)

yの標準偏差 $= |a| \times$ (xの標準偏差)