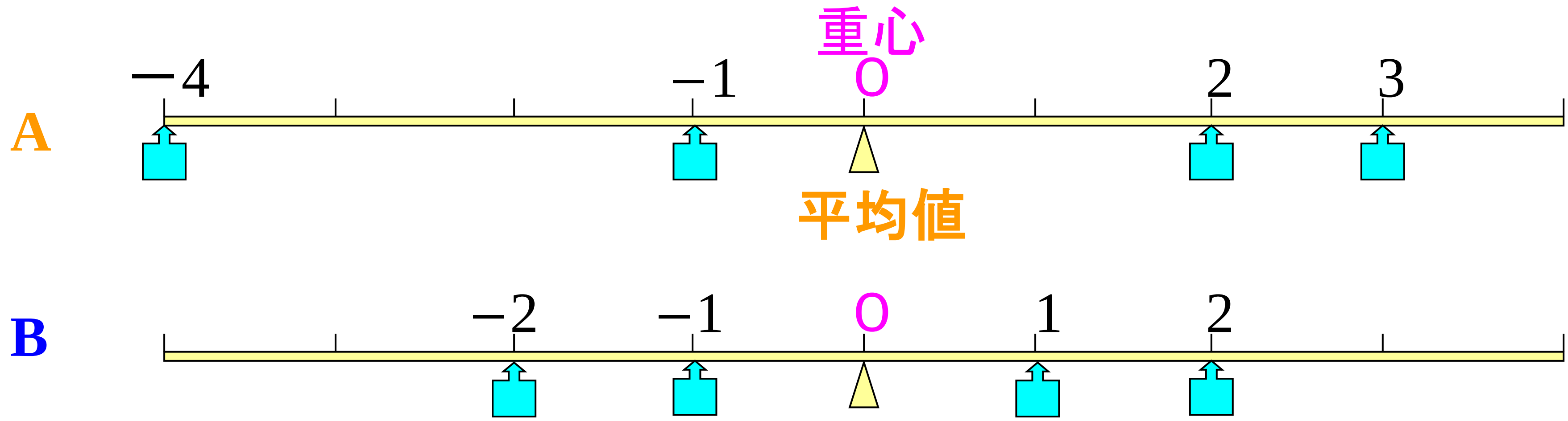
ばらつきを数値化する



天秤の中心から各々上図のような距離に分銅がぶらさがり釣り合っている場合を考える。

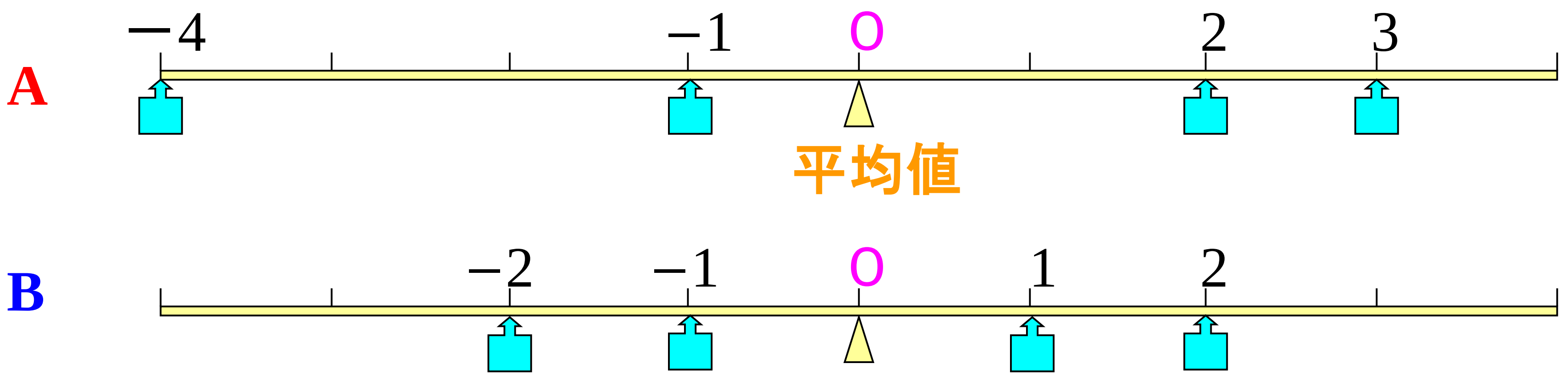
天秤の重心から分銅までの距離は

Aの場合：−4，−1，2，3　　　Bの場合：−2，−1，1，2

距離の平均を求めると、いずれも0になりばらつきの程度を表すことは不可

Aの場合：｛(−4)＋(−1)＋2＋3｝／4＝0

Bの場合：｛(−2)＋(−1)＋1＋2｝／4＝0



そこで、(**距離**)^2＝(**数値ー平均値**)^{2}の**平均を求める**ことにすると、

Aの場合: $\{(-4-0)^2+(-1-0)^2+(2-0)^2+(3-0)^2\}/4$

$=\{(-4)^2+(-1)^2+2^2+3^2\}/4=\mathbf{7.5}$

Bの場合: $\{(-2-0)^2+(-1-0)^2+(1-0)^2+(2-0)^2\}/4$

$=\{(-2)^2+(-1)^2+1^2+2^2\}/4=\mathbf{2.5}$

Aの数値がBより大きいことより

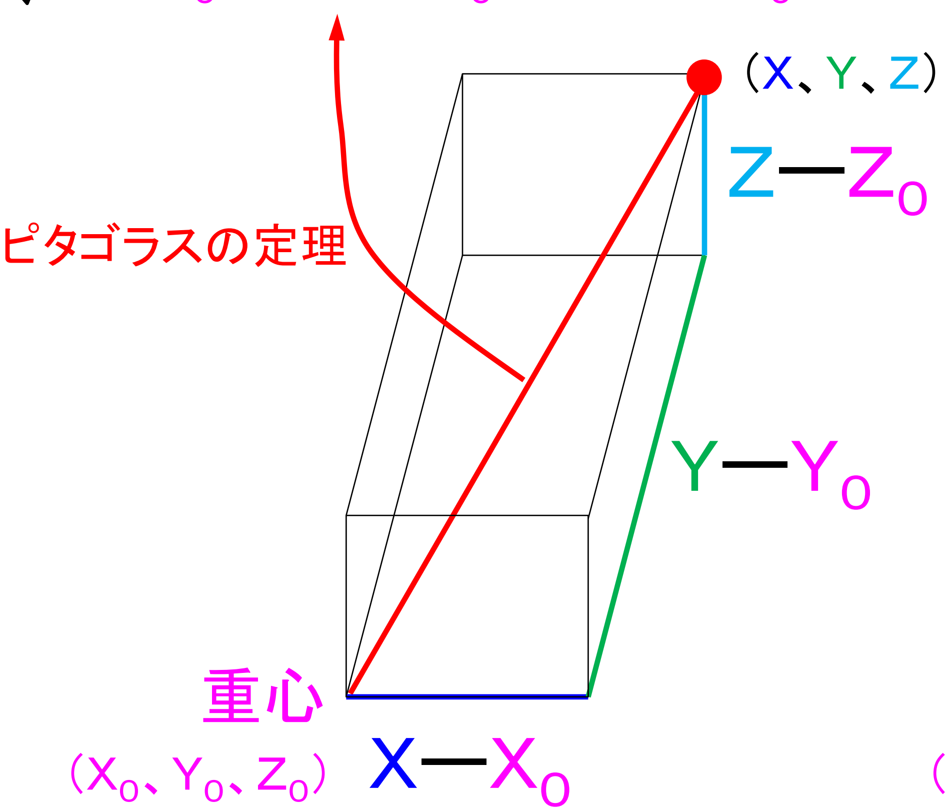
AがBよりばらつきが大きい

●の重心からのばらつきは距離に比例しそうである

（X_0、Y_0、Z_0）

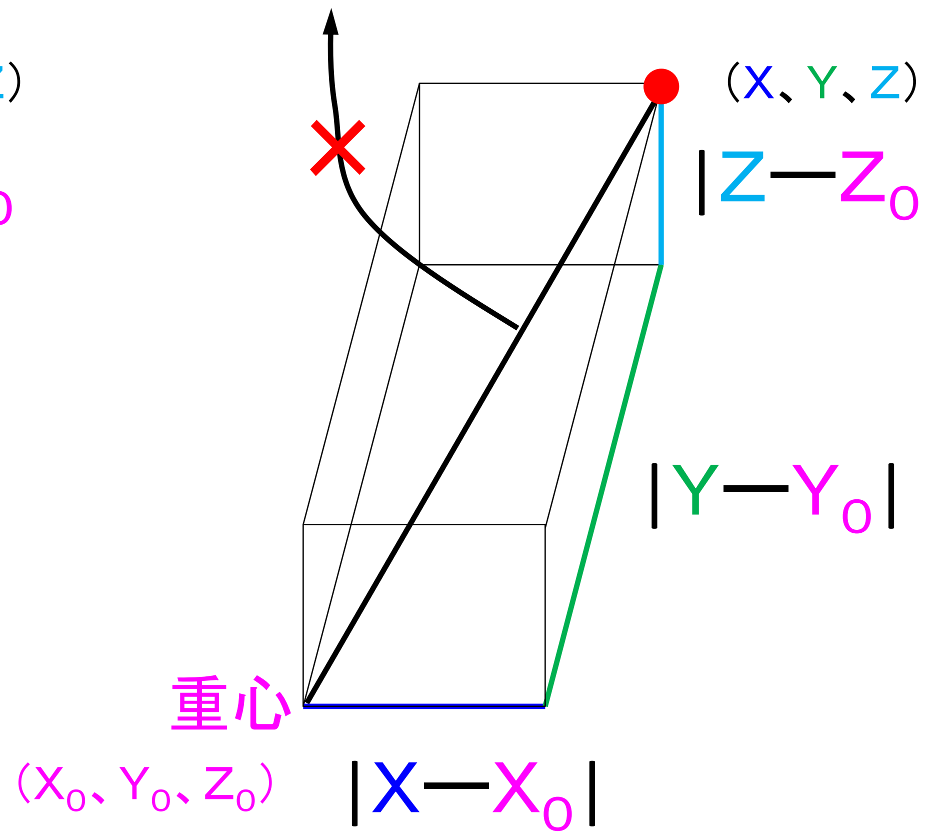
ユークリッド距離

$$\sqrt{(X-X_0)^2+(Y-Y_0)^2+(Z-Z_0)^2}$$

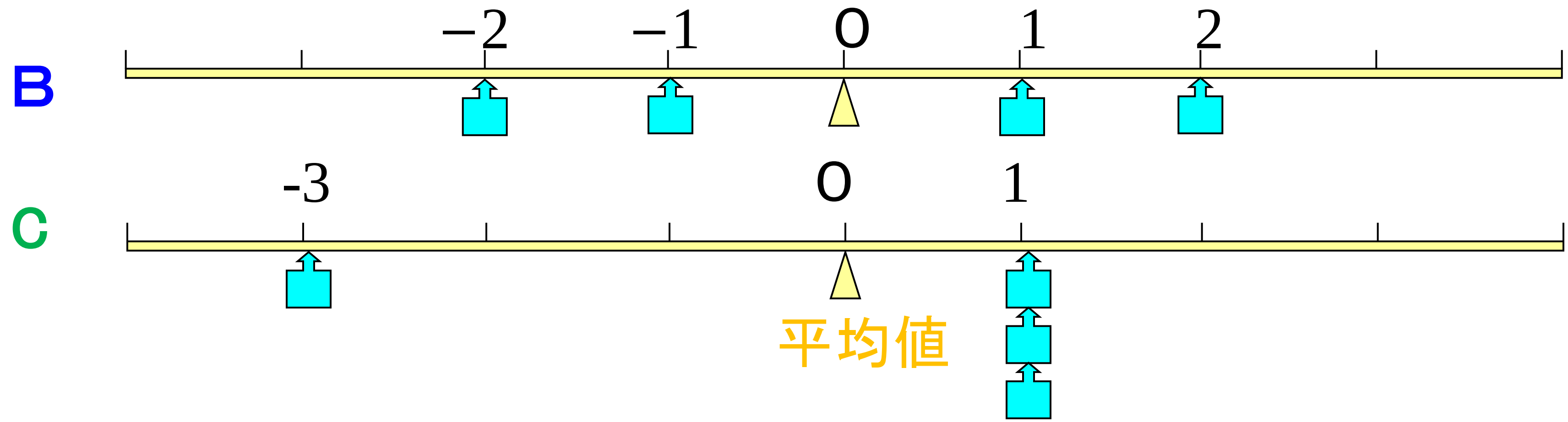


マンハッタン距離

$$|X-X_0|+|Y-Y_0|+|Z-Z_0|$$



ばらつきを数値化する（おまけ）



（**距離**）^{2}でなくて、絶対値の和ではなぜいけないか？

Aの場合：$(|-4|+|-1|+|2|+|3|)/4=2.5$

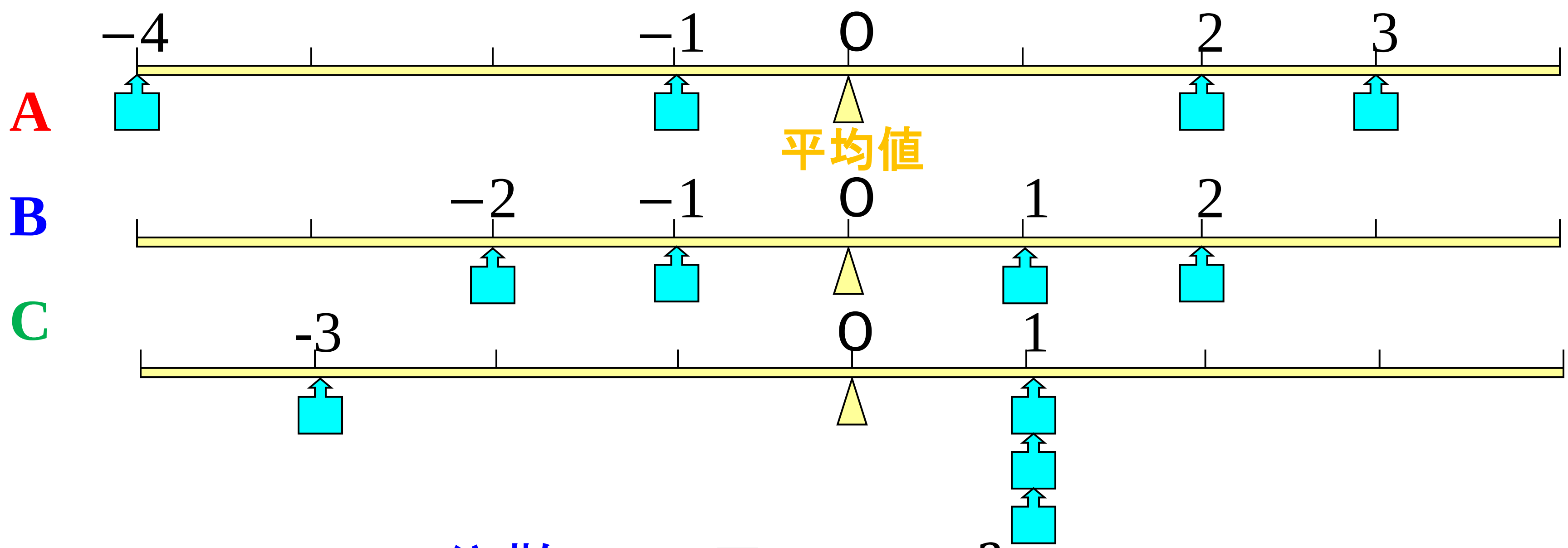
Bの場合：$(|-2|+|-1|+|1|+|2|)/4=1.5$

C絶対値：$\left(\overbrace{|-3|}^{3}+\overbrace{|1|+|1|+|1|}^{3}\right)/4=1.5$

同じ値

C平方和：$\left\{\overbrace{(-3)^2}^{9}+\overbrace{1^2+1^2+1^2}^{3}\right\}/4=3.0$

平方和の方が、平均値から離れた距離の重さ（パワー）が反映されるから



ばらつきの程度→分散 $V=\sum(x-\bar{x})^2/n$　平方和の平均

Aの場合：$\{(-4-0)^2+(-1-0)^2+(2-0)^2+(3-0)^2\}/4=7.5$　ばらつき大

Bの場合：$\{(-2-0)^2+(-1-0)^2+(1-0)^2+(2-0)^2\}/4=2.5$

C平方和：$\{(-3-0)^2+(1-0)^2+(1-0)^2+(1-0)^2\}/4=3.0$

標準偏差 $\sigma=\sqrt{分散V}=\sqrt{\sum(x-\bar{x})^2/n}$

Aの場合：$\sigma_A=\sqrt{7.5}\ =2.73$　ばらつき大

Bの場合：$\sigma_B=\sqrt{2.5}\ =1.58$

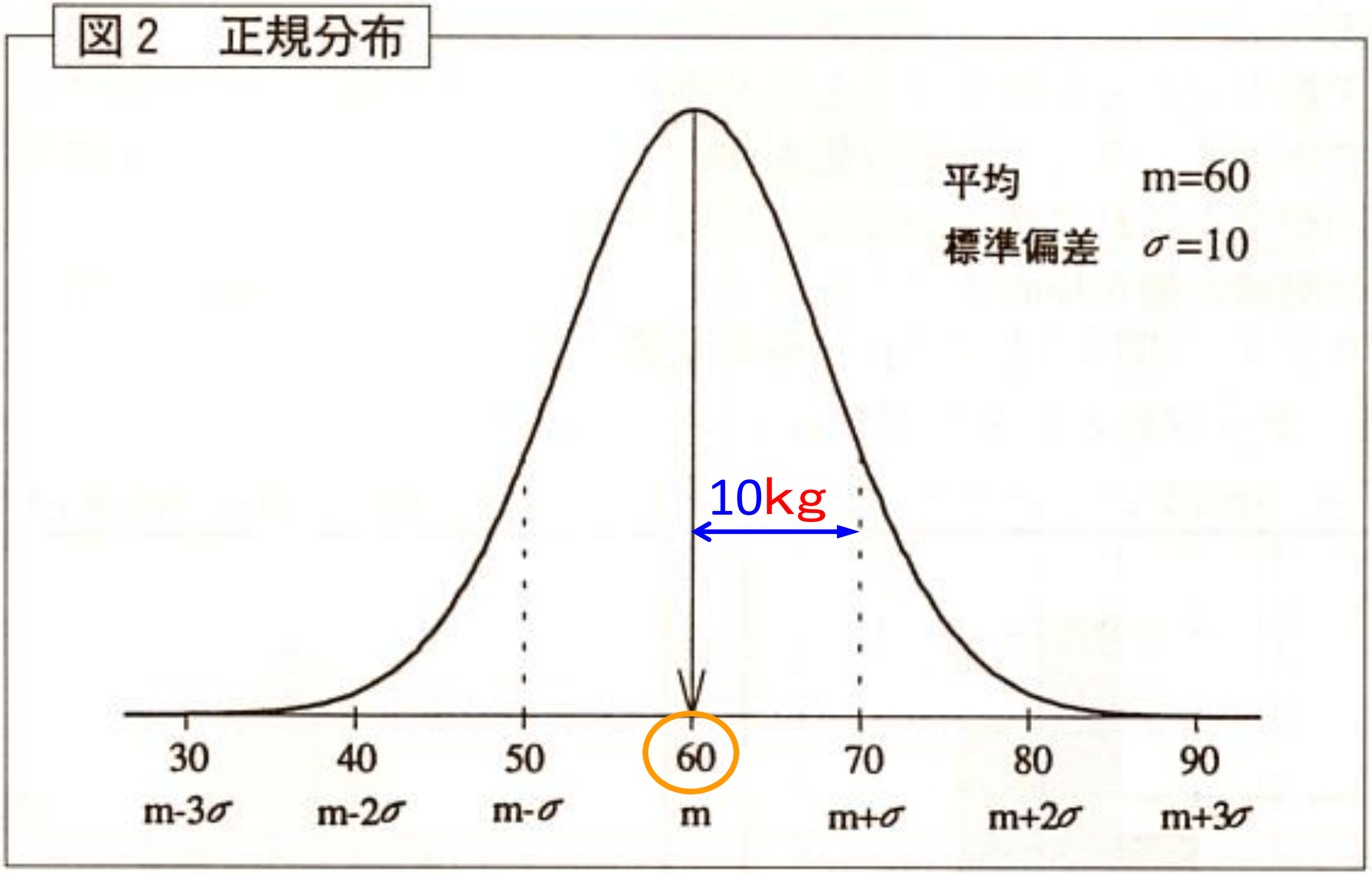
Cの場合：$\sigma_C=\sqrt{3.0}\ =1.73$

$\bar{x}$：平均値

n：数

なぜ平方根にするか？　→　単位（次元）を合わせるため

平均体重が60kg、標準偏差σが10kgの分布
（分散Vは100kg²）



なぜ平方根にするか？ → 単位：　kg