統計演習１　中央極限定理

平均すると、何でも正規分布に収まる

統計演習２　T検定

多数集めると差が見える

統計演習３　最小二乗法

一番フィットする関数は

統計演習１　中央極限定理

平均すると、何でも正規分布に収まる

統計演習２　T検定

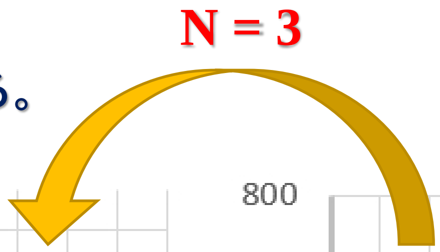
多数集めると差が見える

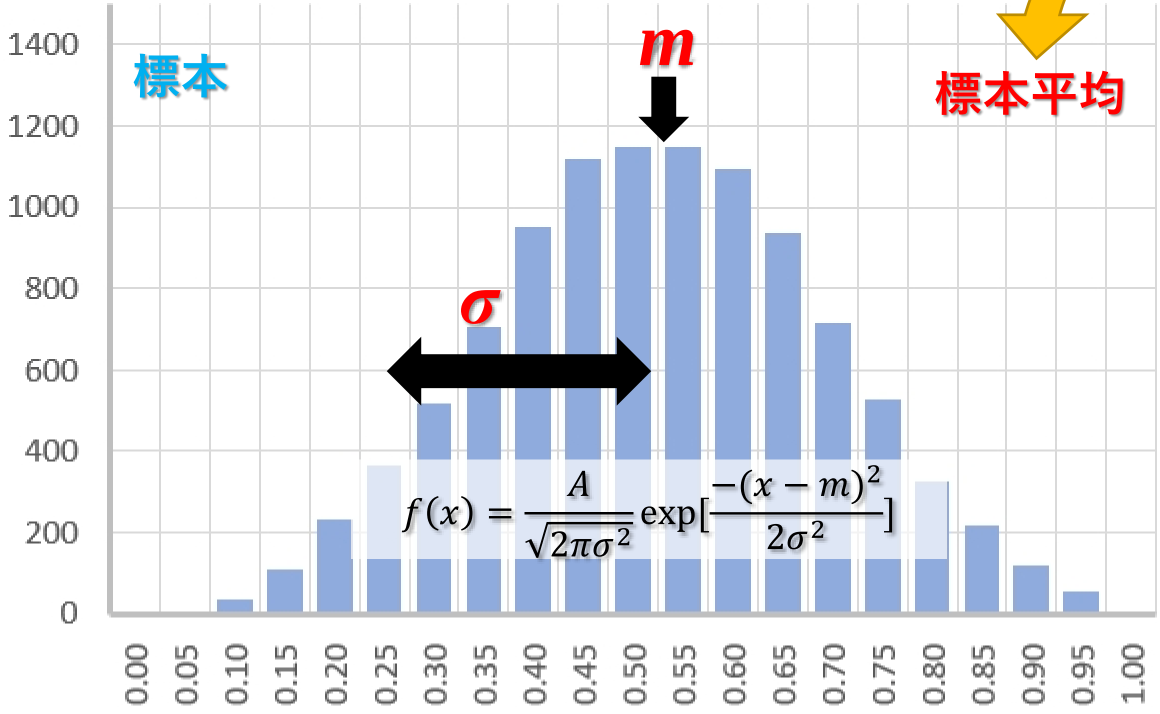
統計演習３　最小二乗法

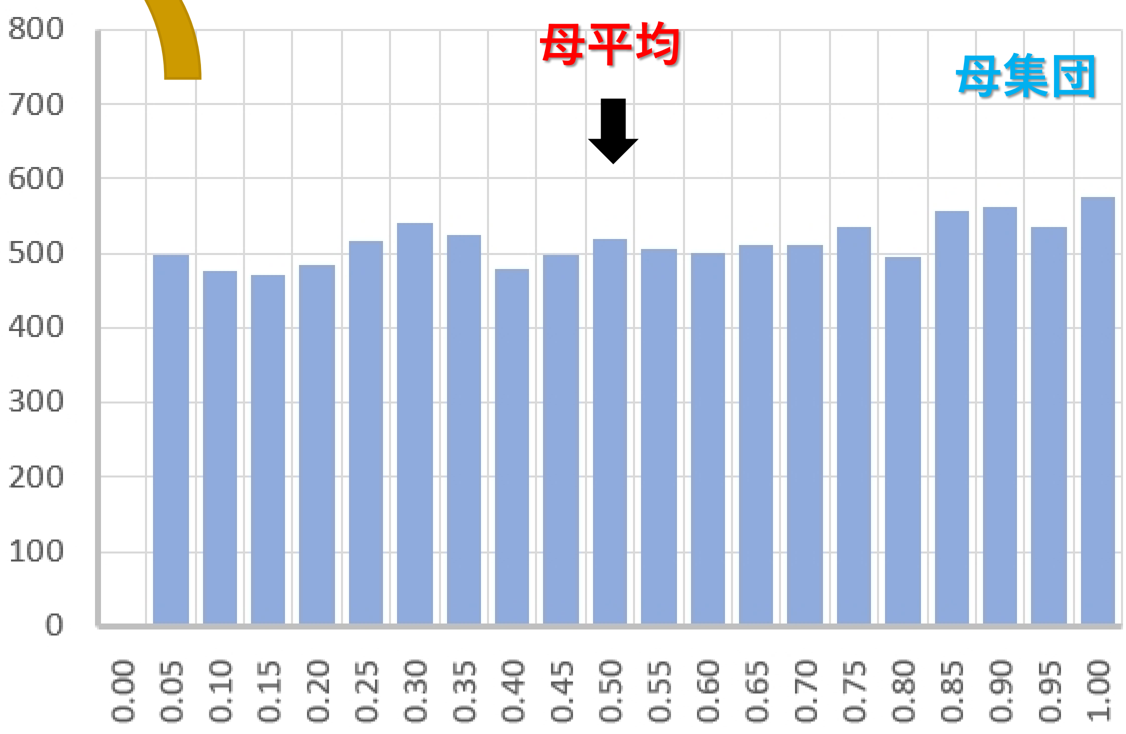
一番フィットする関数は

中央極限定理

もとの分布がどのようなものであっても、
ランダムに抽出して求めた平均値は正規分布となる。







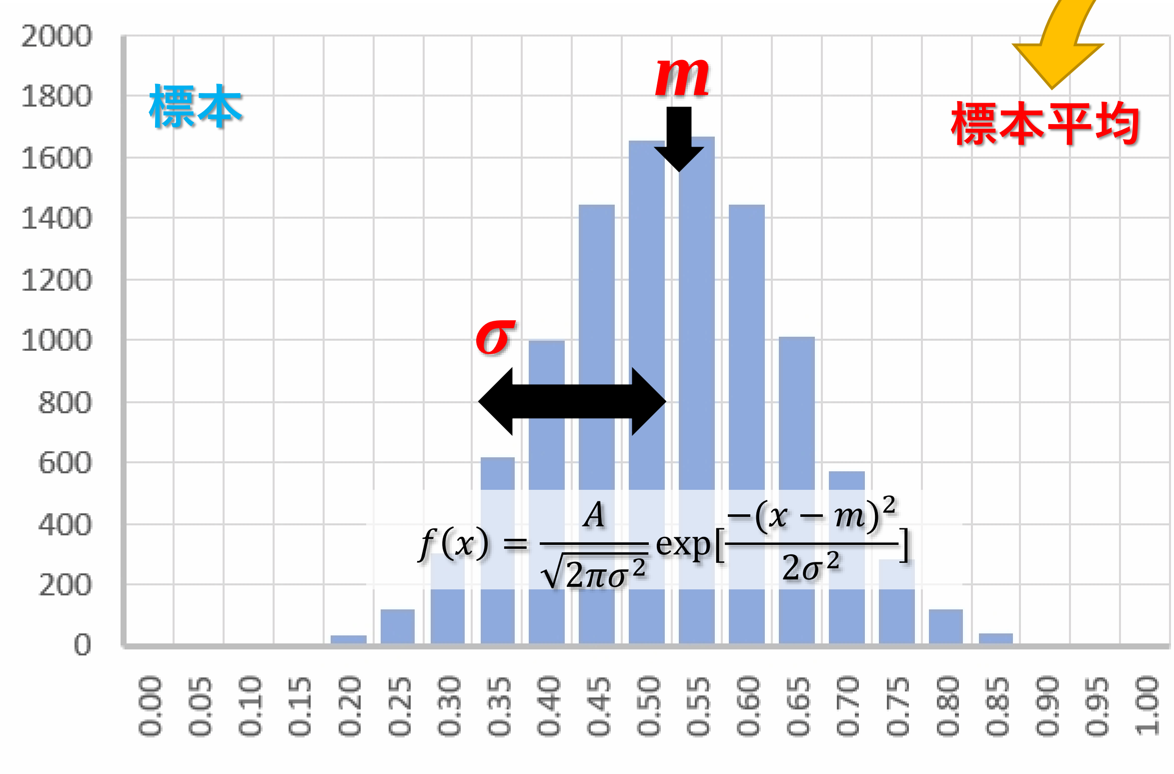
任意抽出したデータの平均値は正規分布

ランダムな数値

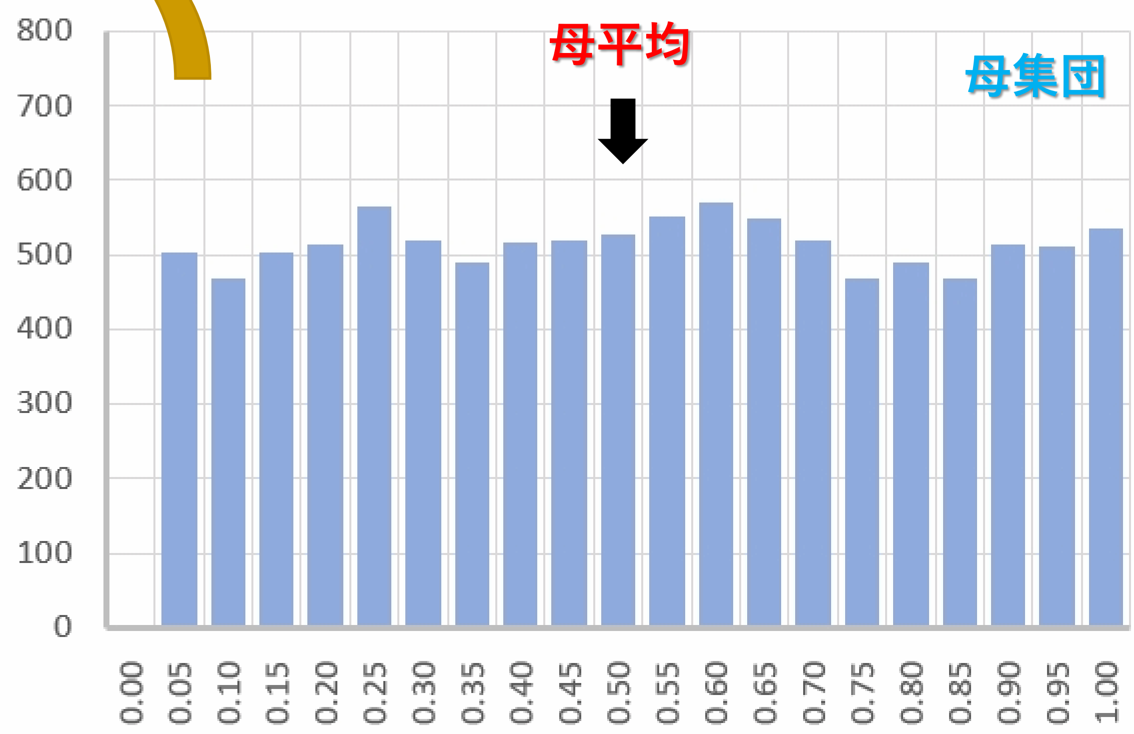
- ●データ数が少ないと平均値の偏差は大きい
- ●m ± σの中に68.2%（約７割）のデータを含む
- ●母平均 ≠ 標本平均となるのが通常

中央極限定理

もとの分布がどのようなものであっても、
ランダムに抽出して求めた平均値は正規分布となる。



N = 6



任意抽出したデータの平均値は正規分布

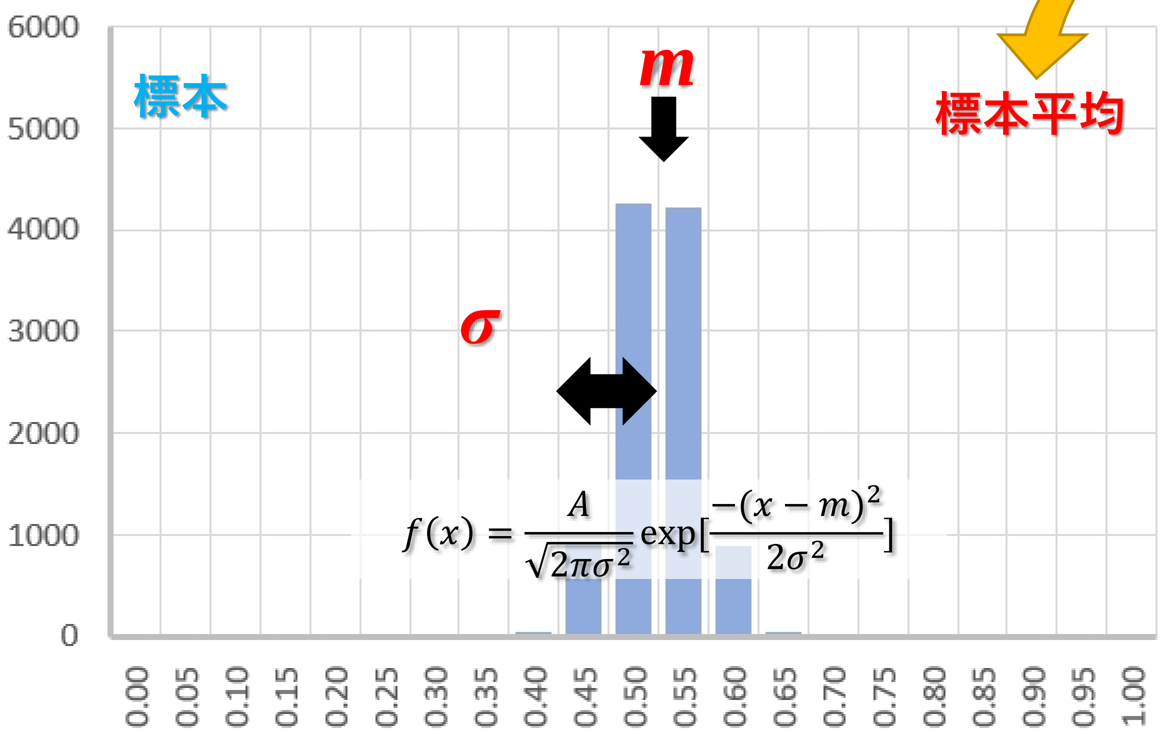
ランダムな数値

●データ数が中程度ならば、中程度の偏差

中央極限定理

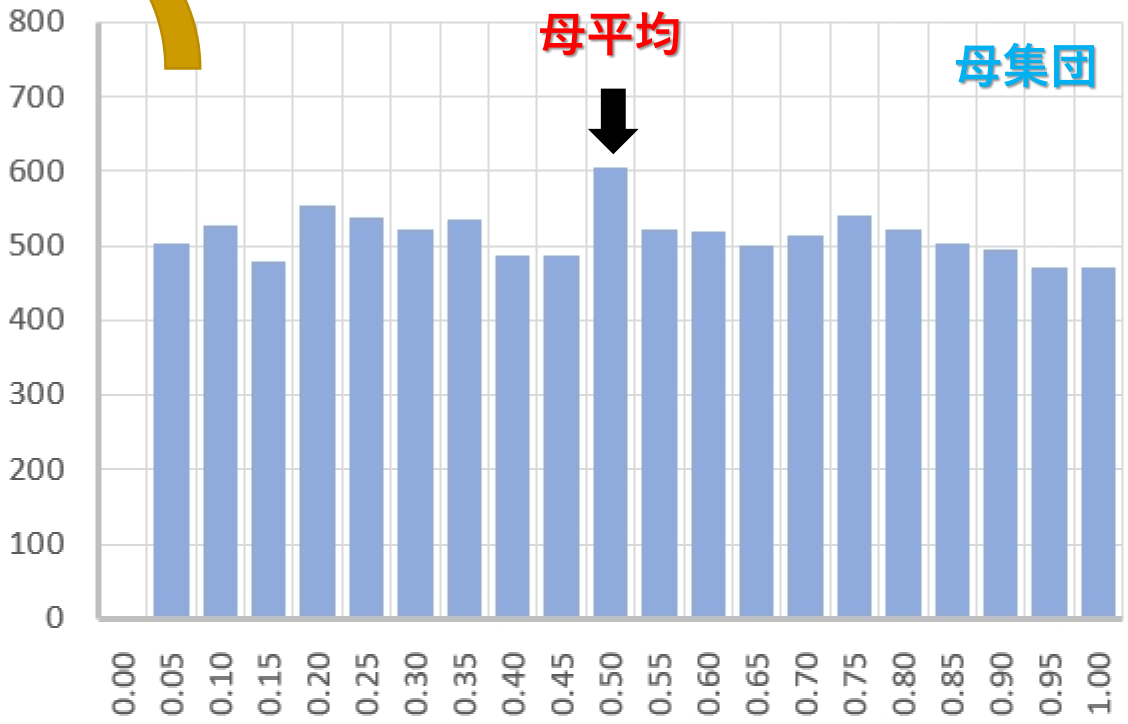
もとの分布がどのようなものであっても、
ランダムに抽出して求めた平均値は正規分布となる。

N = 60



$$f(x) = \frac{A}{\sqrt{2\pi\sigma^2}}\exp[\frac{-(x-m)^2}{2\sigma^2}]$$

任意抽出したデータの平均値は正規分布

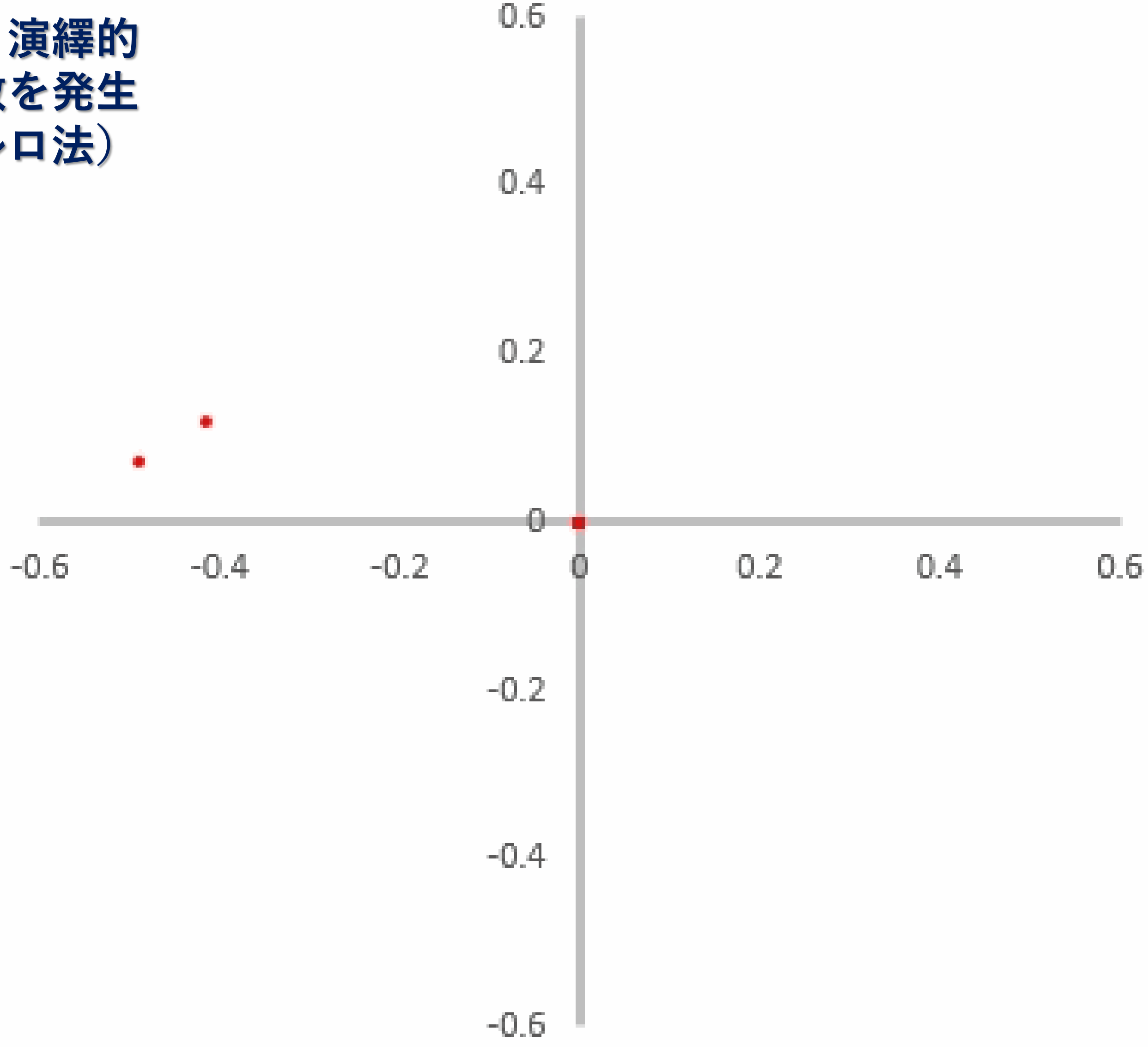


ランダムな数値

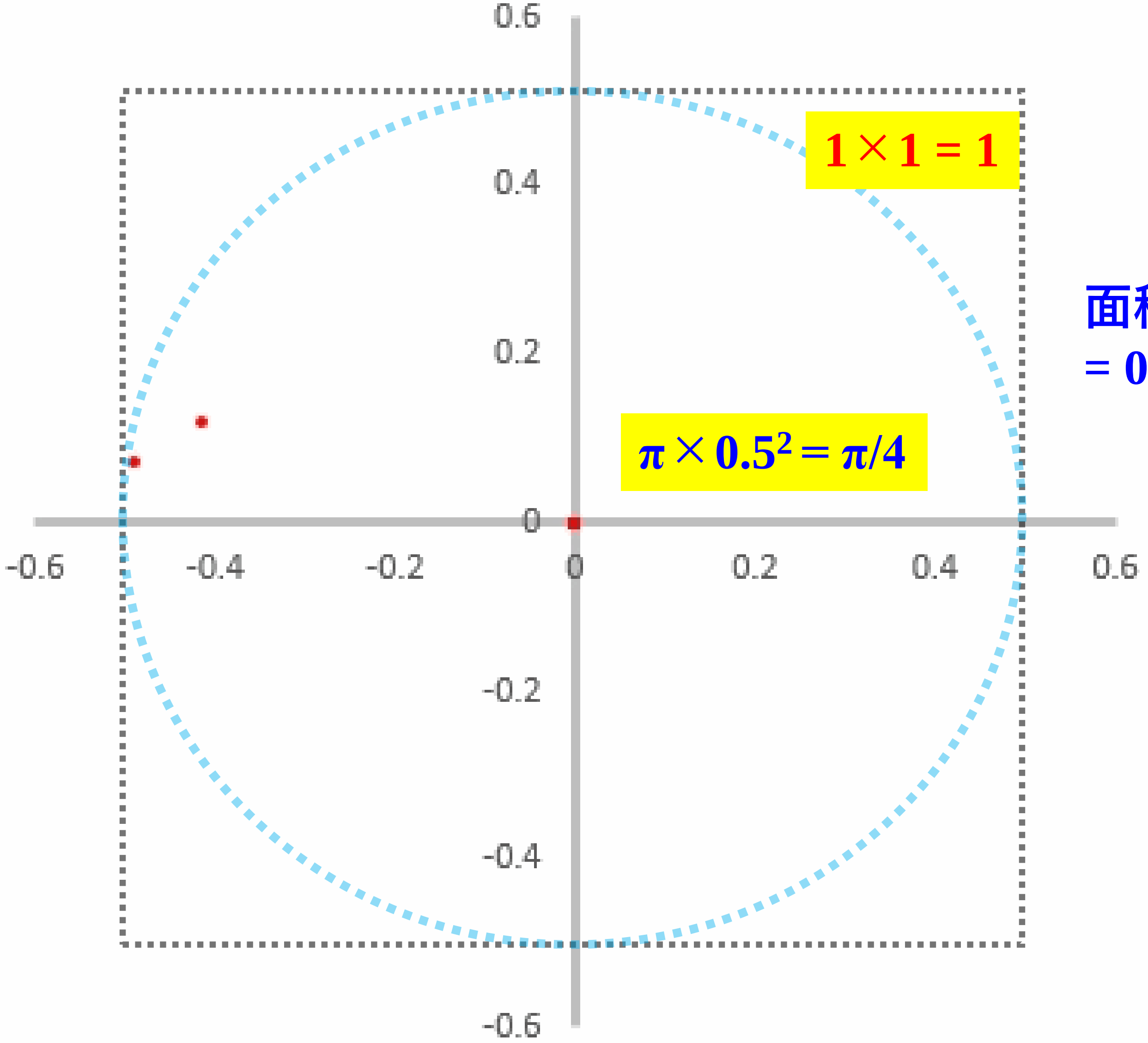
●データ数が多いと、平均値（標本平均）の偏りは小さく、より正確に元の平均値（母平均）を反映

乱数計算の応用

明確な関数や理論によって、演繹的に予測できない現象を、乱数を発生させて予測する（モンテカルロ法）



課題：中心から0.5（単位なし）の距離以内の点の数を数える

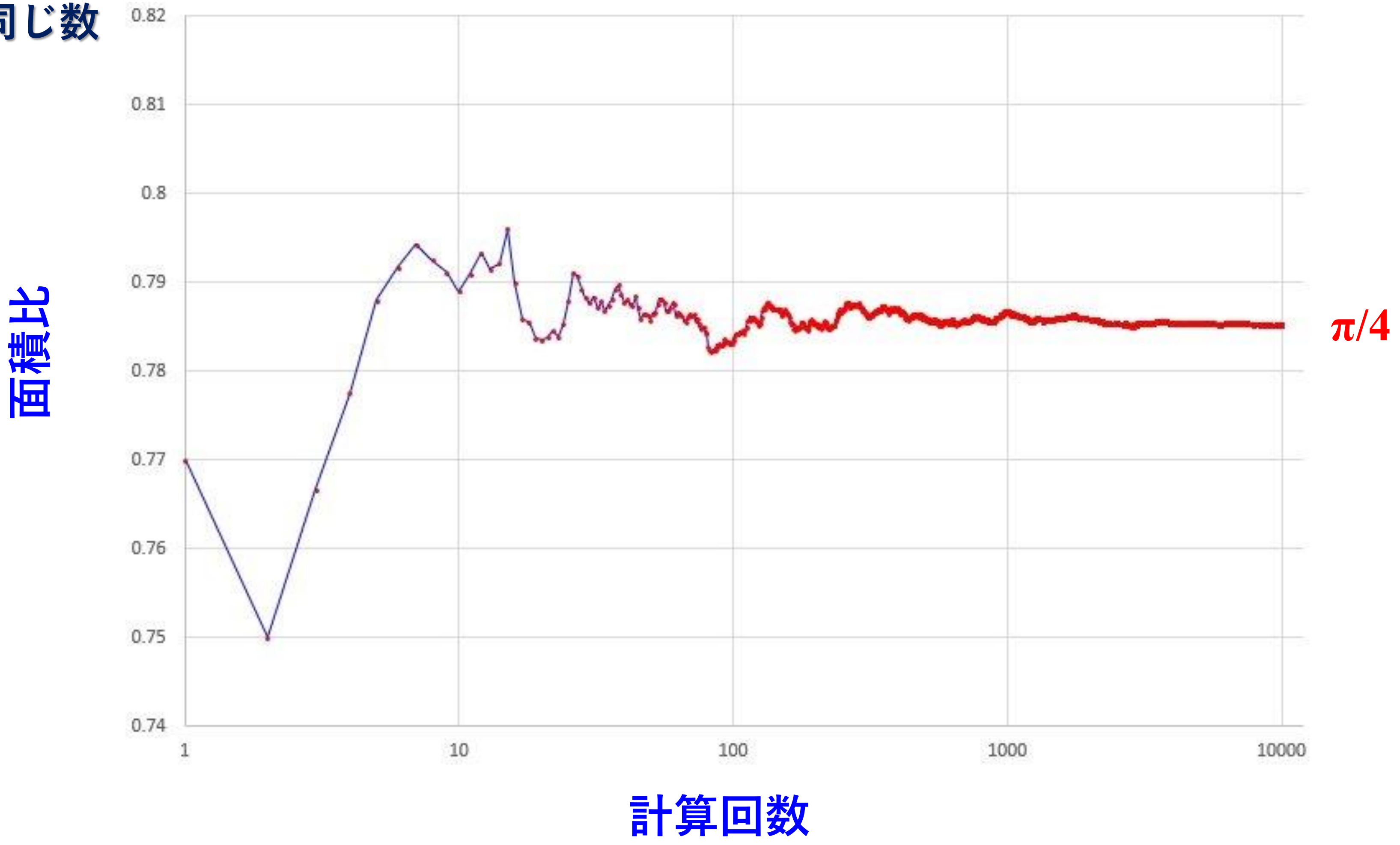


面積比 = $\pi/4$

= 0.785398163397448・・・・

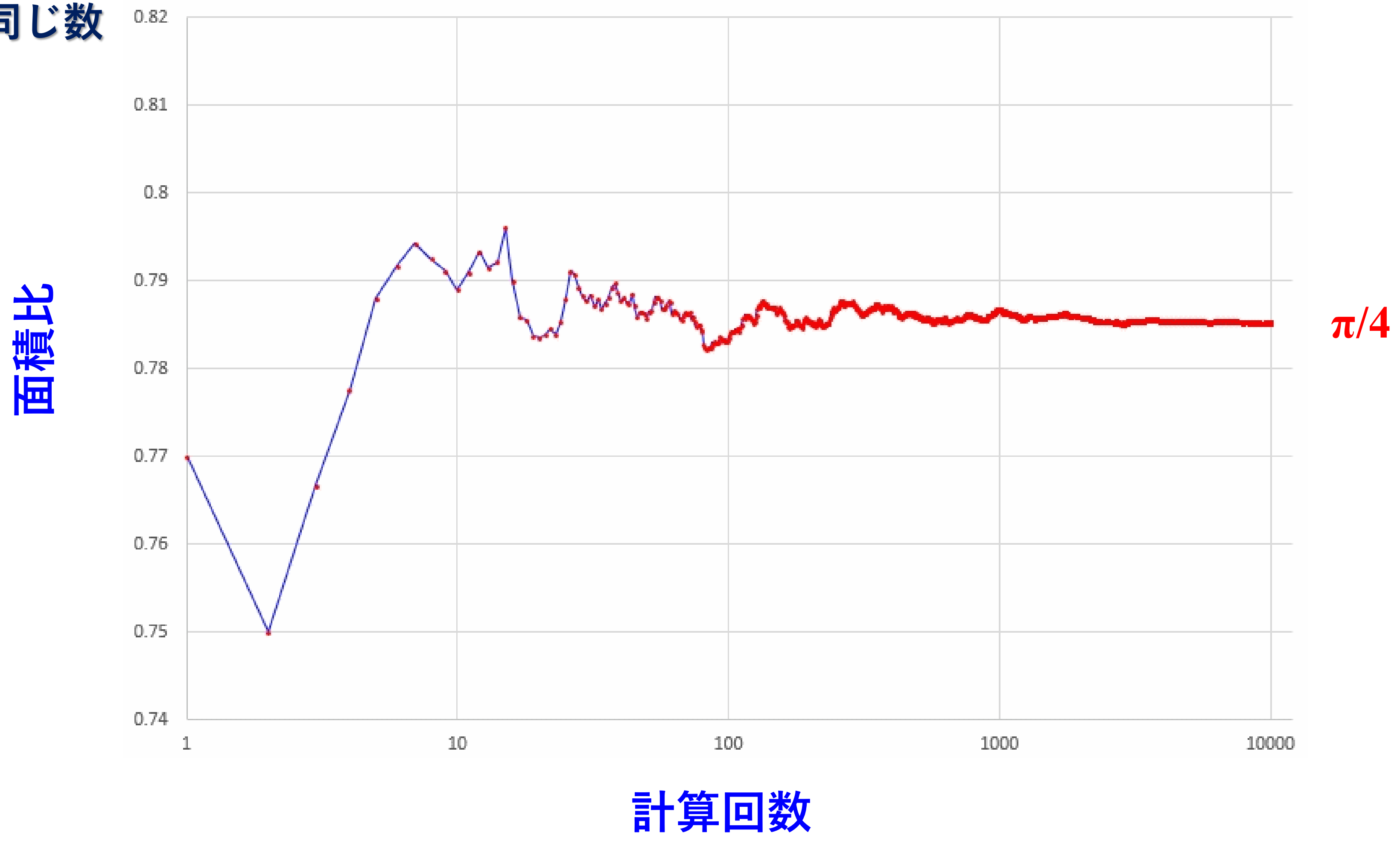
モンテカルロ法

計算ごとに結果は少しずつ異なるが、同じ数値に近づく。



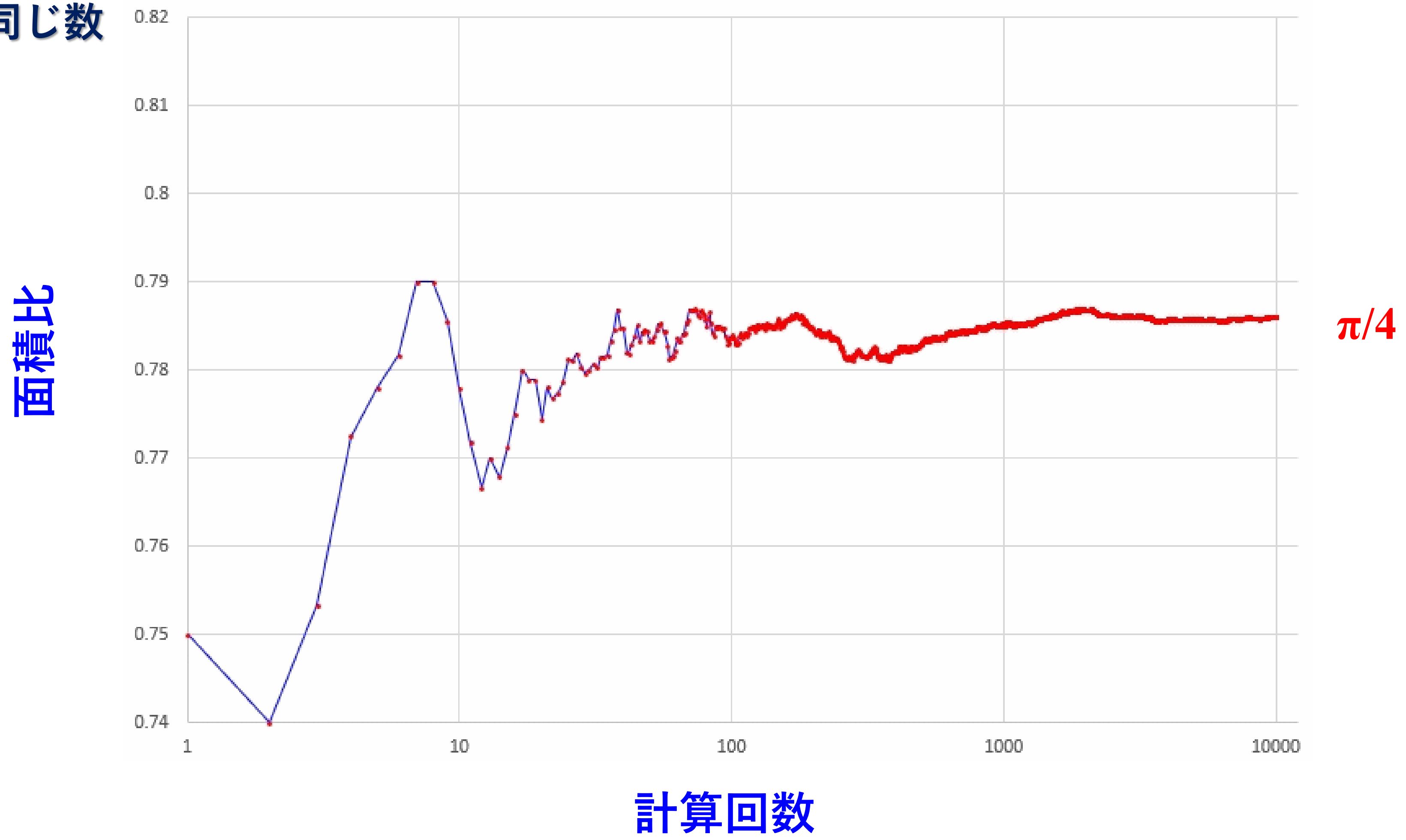
モンテカルロ法

計算ごとに結果は少しずつ異なるが、同じ数値に近づく。



モンテカルロ法

計算ごとに結果は少しずつ異なるが、同じ数値に近づく。



統計演習１　中央極限定理

平均すると、何でも正規分布に収まる

統計演習２　T検定

多数集めると差が見える

統計演習３　最小二乗法

一番フィットする関数は

T検定

命題：２つのグループの母平均は異ならないか？

↓

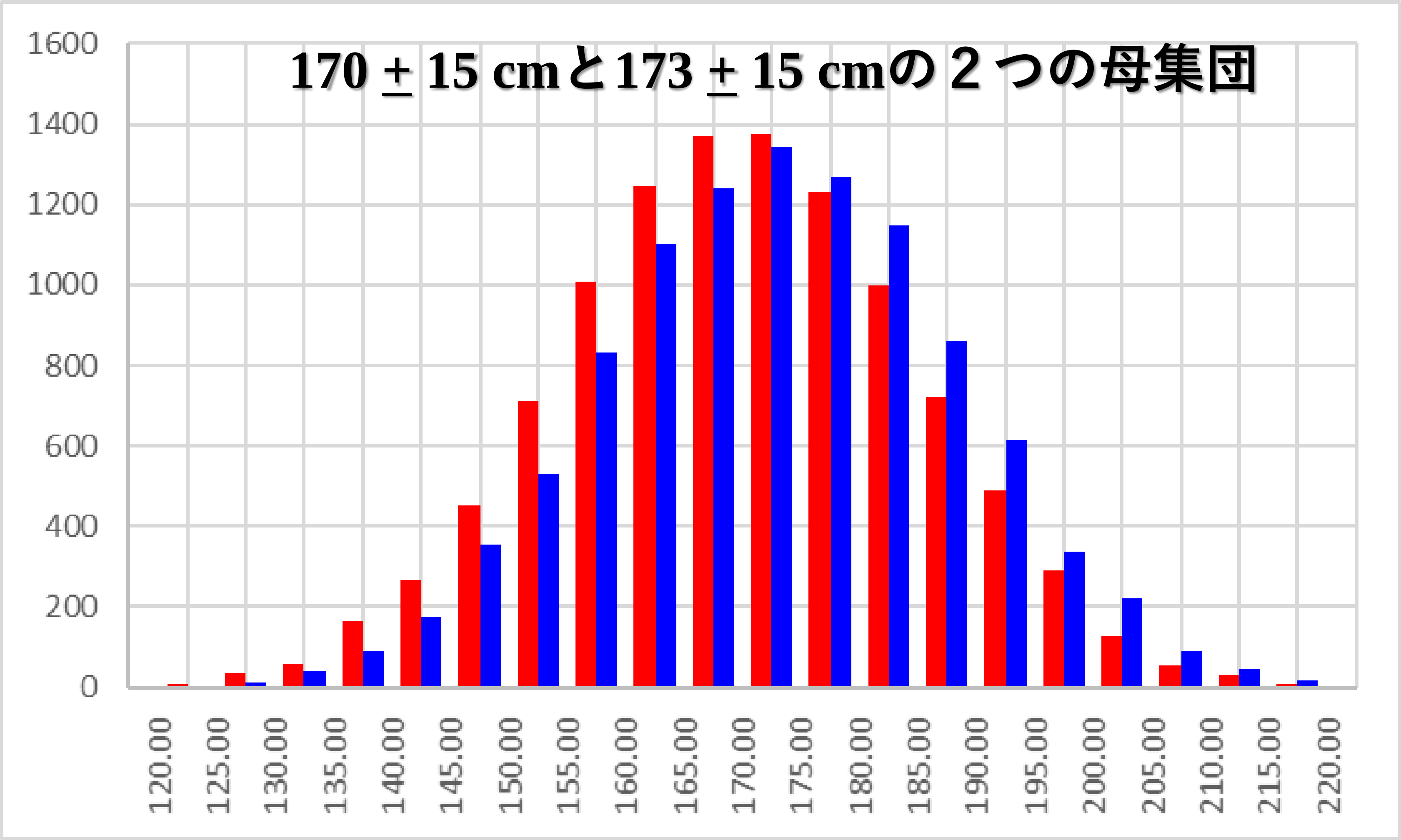
帰無仮説を否定できる確率　*p*

一般にP < 0.01〜0.001の場合、元の2つの集団の平均は同じではないと言う。

↓

有意差がある

‘Significant difference’



標本数（n）によって示せる有意差が異なる。

p = 0.254 ± 0.140 (n=3, p = 0.00437〜0.492)

p = 0.210 ± 0.137 (n=30, p = 0.00107〜0.494)

p = 0.0468 ± 0.101 (n=300, p = 0.00000204〜0.482)

T検定

命題：２つのグループの母平均は異ならないか？

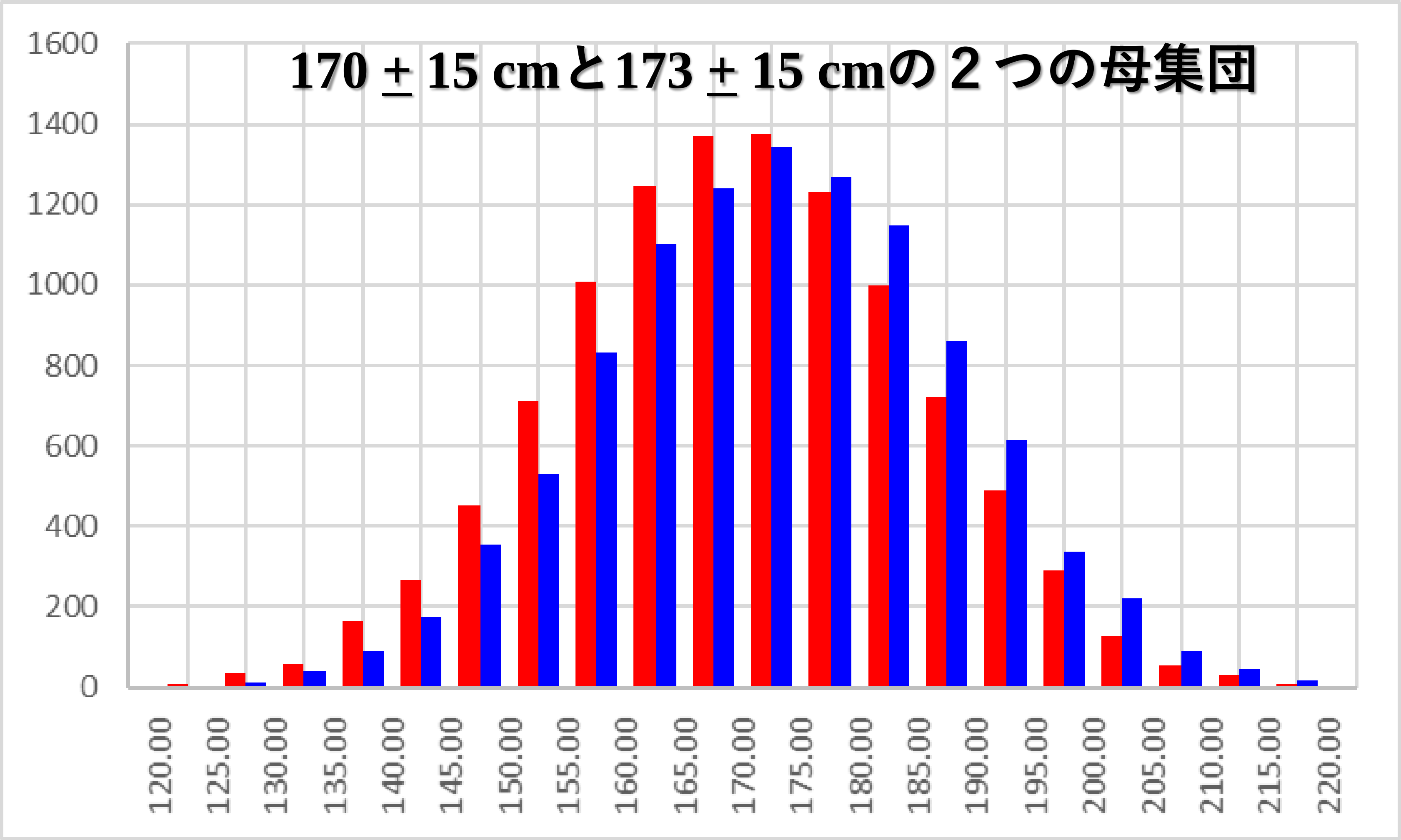
↓

帰無仮説を否定できる確率　*p*

一般に P < 0.01～0.001 の場合、元の2つの集団の平均は同じではないと言う。

↓

有意差がある
'Significant difference'



標本数（n）によって示せる有意差が異なる。

・n～3程度では、有意差を示すことは難しい。
・n～30程度で、有意差を示せることがある。
・n～300では、有意差を示すことが可能である。

T検定

命題：２つのグループの母平均は異ならないか？

↓

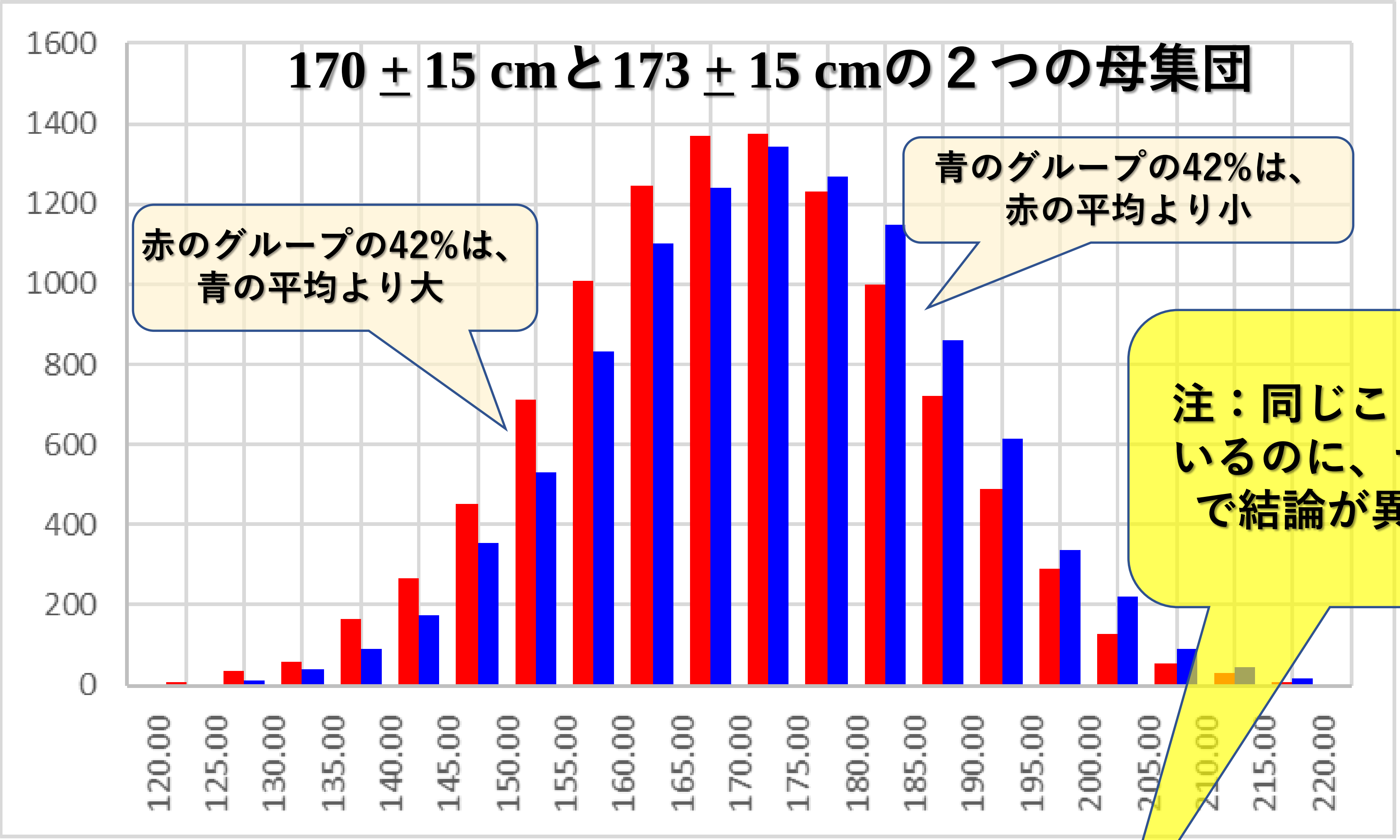
帰無仮説を否定できる確率　*p*

一般に P < 0.01～0.001 **の場合、元の2つの集団の平均は同じではないと言う。**

↓

有意差がある

‘Significant difference’



注：同じことをしているのに、データ数で結論が異なる。

標本数（n）によって示せる有意差が異なる。

・n～3程度では、有意差を示すことは難しい。
・n～30程度で、有意差を示せることがある。
・n～300では、有意差を示すことが可能である。

統計演習１　中央極限定理

平均すると、何でも正規分布に収まる

統計演習２　T検定

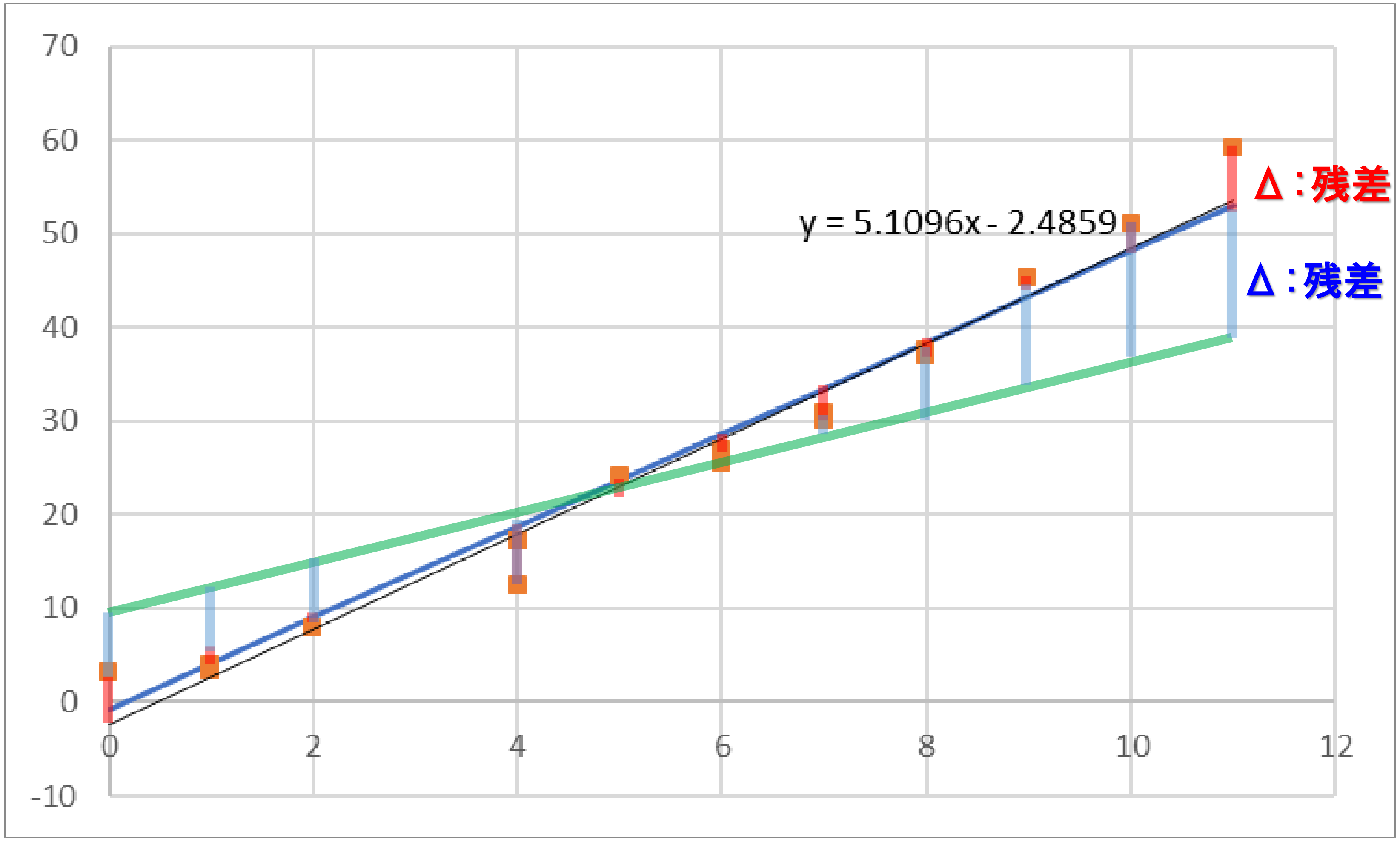
多数集めると差が見える

統計演習３　最小二乗法

一番フィットする関数は

最小二乗法による直線回帰（Linear regression by least square method）

従属変数（y）

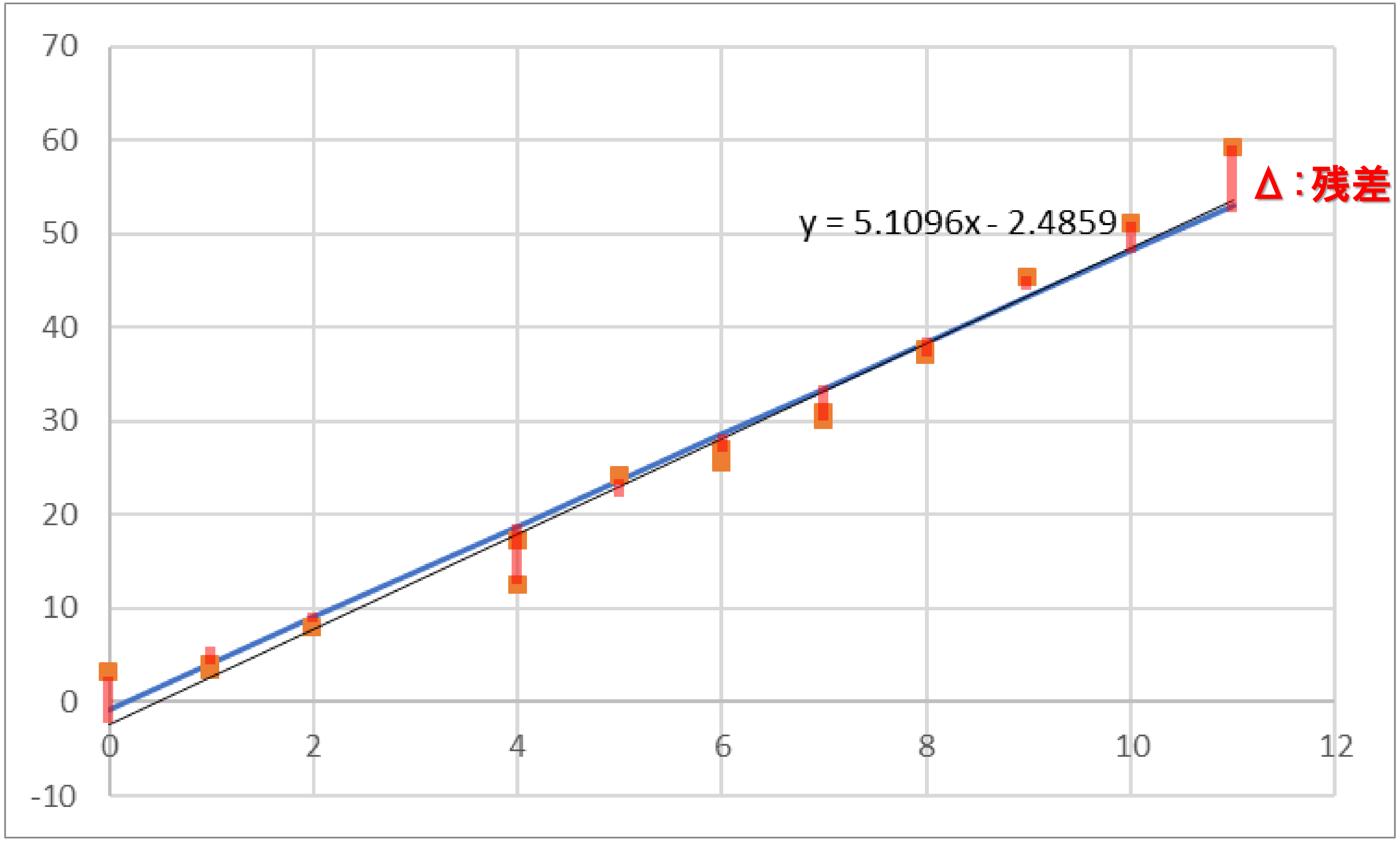


Δ：残差 → ΣΔ²を最小とするaとb

Δ：残差

独立変数（x）

最小二乗法による直線回帰（Linear regression by least square method）

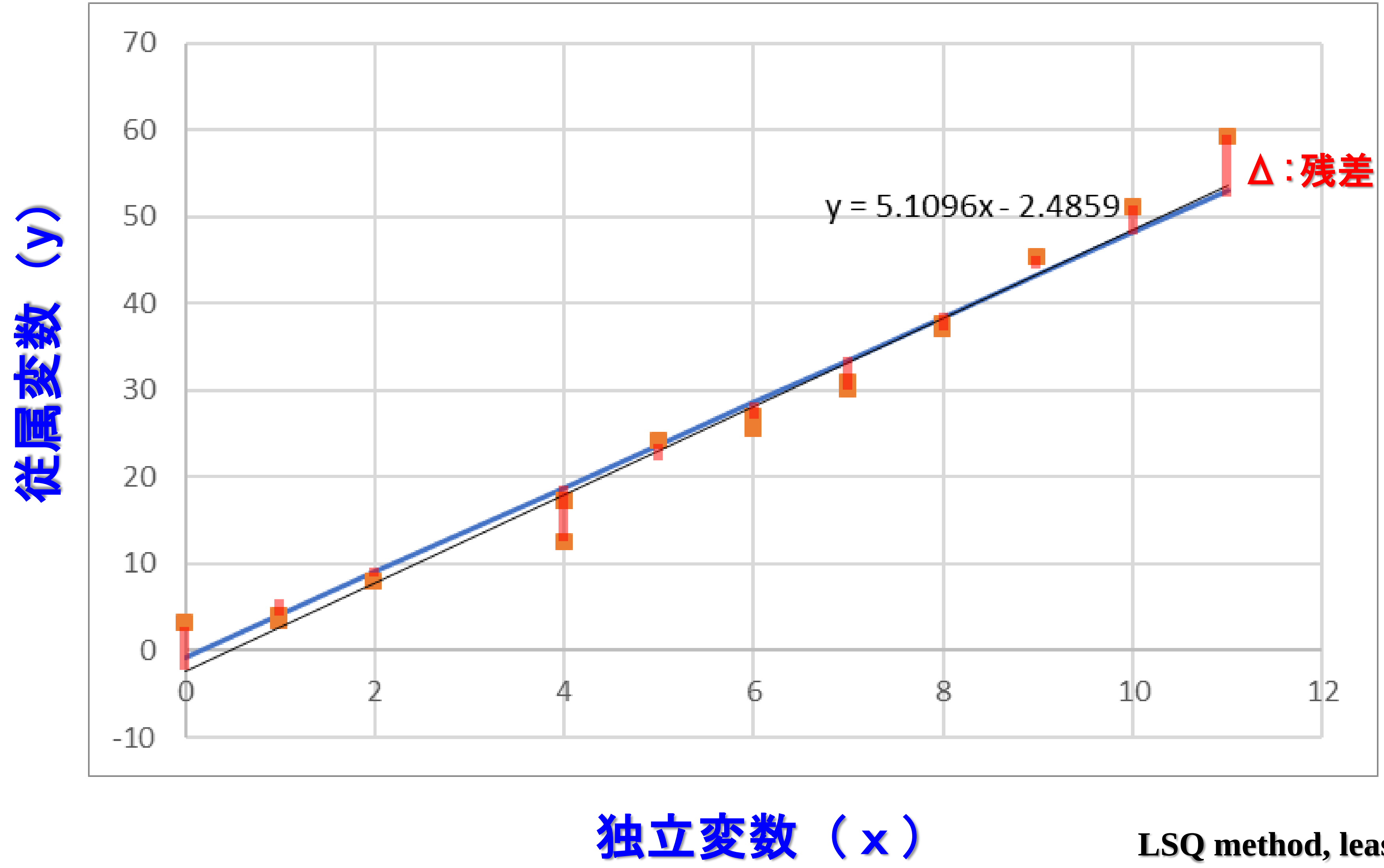


従属変数（y）

独立変数（x）

Δ：残差 → ΣΔ²を最小とするaとb

最小二乗法による直線回帰（Linear regression by least square method）



Δ：残差 → ΣΔ^{2}を最小とするaとb

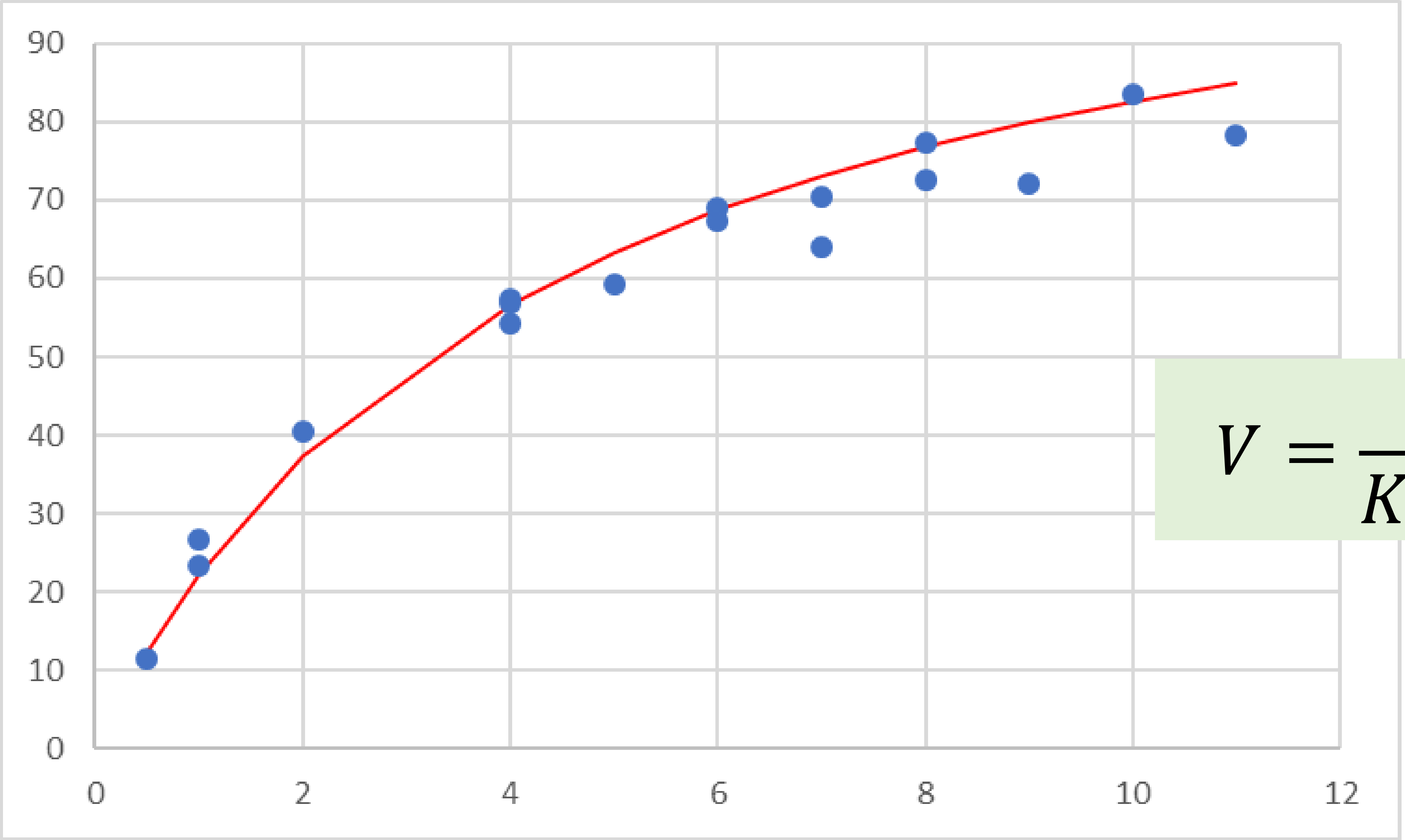
$$a = \frac{s_{xy}}{{s_x}^2}$$

$$= \frac{\sum_{n=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{n=1}^{n}(x_i - \overline{x})^2}$$

$$b = \overline{y} - a\overline{x}$$

LSQ method, least square method, linear regression, linear approximation

最小二乗法による曲線回帰（Curve fitting by least square method）



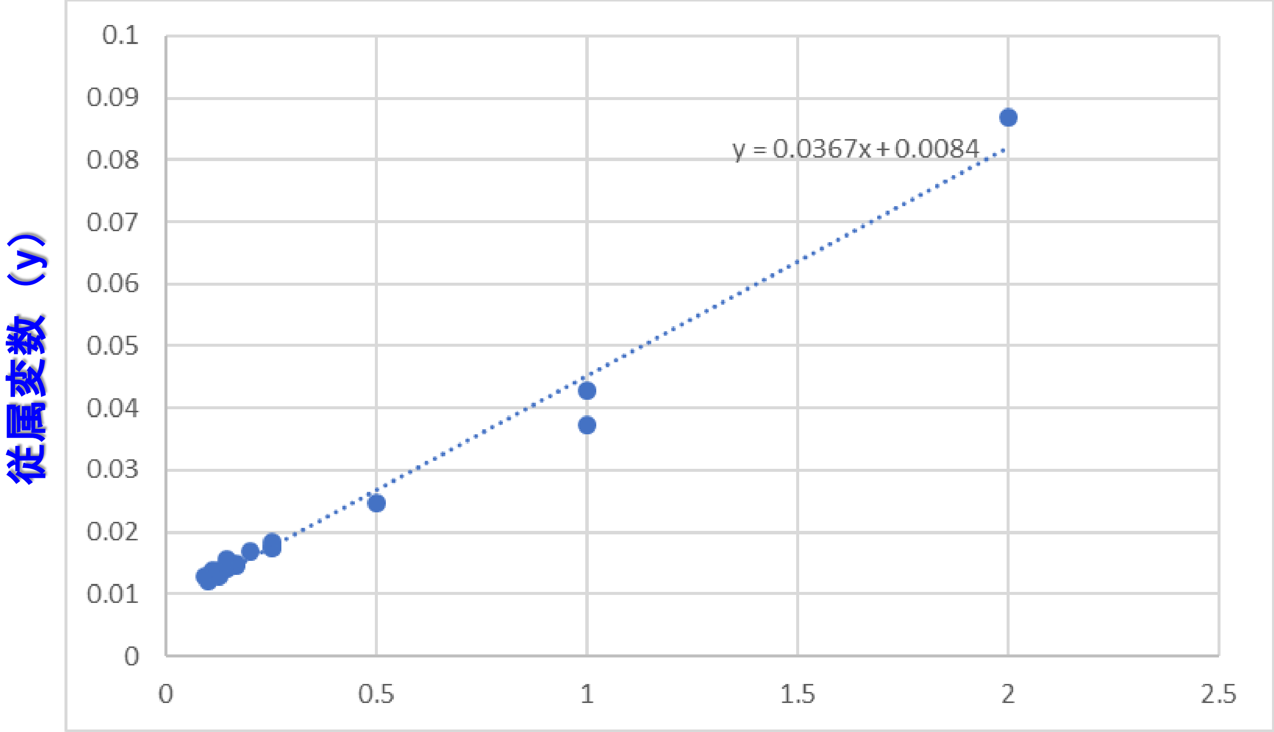
従属変数（y）

$$V = \frac{V_m x}{K_m + x}$$

独立変数（x）

両対数プロットからの直線回帰

$K_m = 4.3$
$V_m = 120$

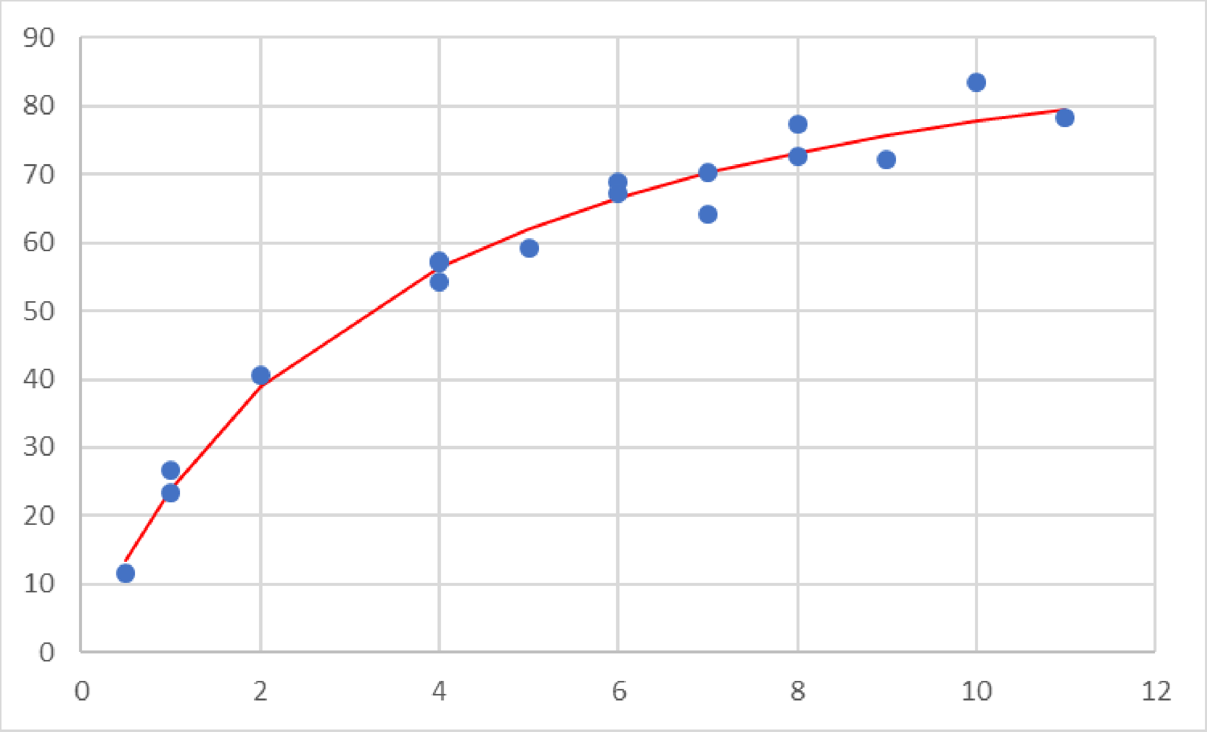


独立変数（x）

$$y = \frac{1}{V} = \frac{1}{V_m} + \frac{K_m}{V_m} \times \frac{1}{[S]}$$
$$= \frac{1}{V_m} + \frac{K_m}{V_m} \times x$$

曲線への最小二乗法による回帰

$K_m = 3.3$
$V_m = 100$

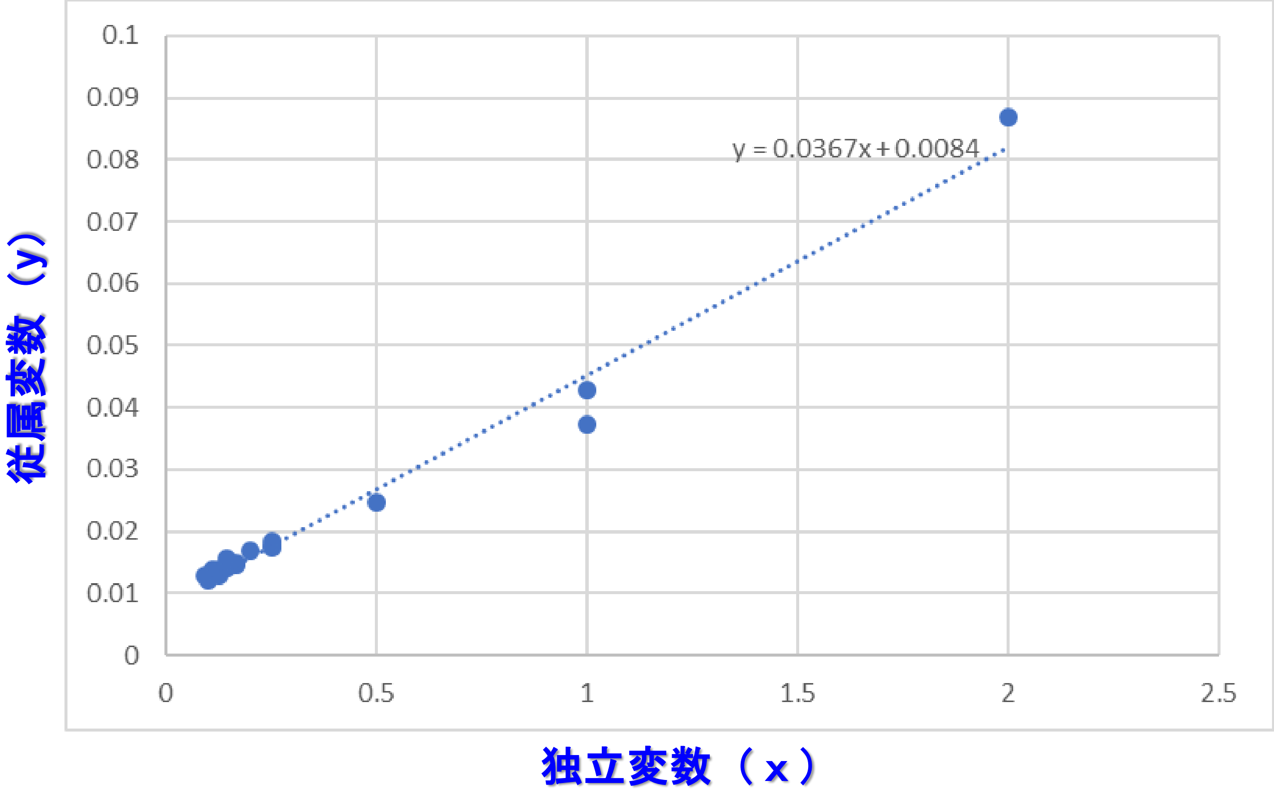


独立変数（x）

$$V = \frac{V_m x}{K_m + x}$$

両対数プロットからの直線回帰

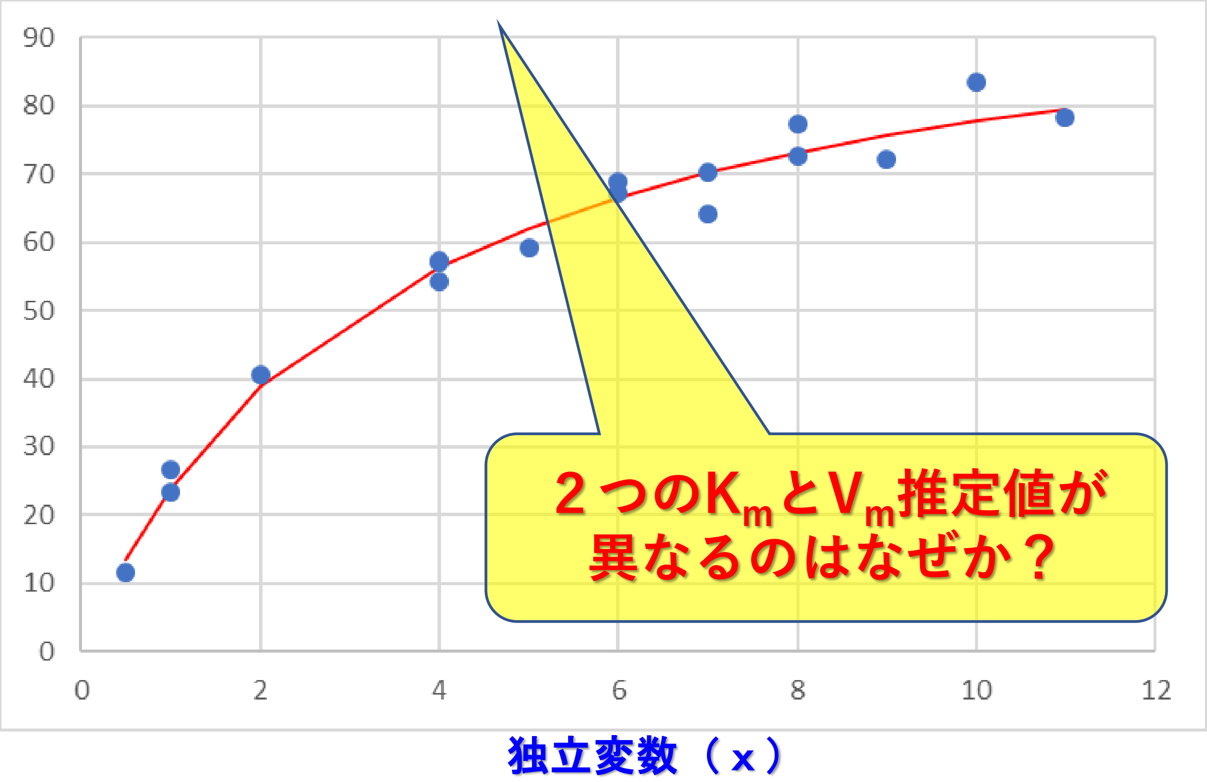
$K_m = 4.3$
$V_m = 120$



$$y = \frac{1}{V} = \frac{1}{V_m} + \frac{K_m}{V_m} \times \frac{1}{[S]}$$
$$= \frac{1}{V_m} + \frac{K_m}{V_m} \times x$$

曲線への最小二乗法による回帰

$K_m = 3.3$
$V_m = 100$

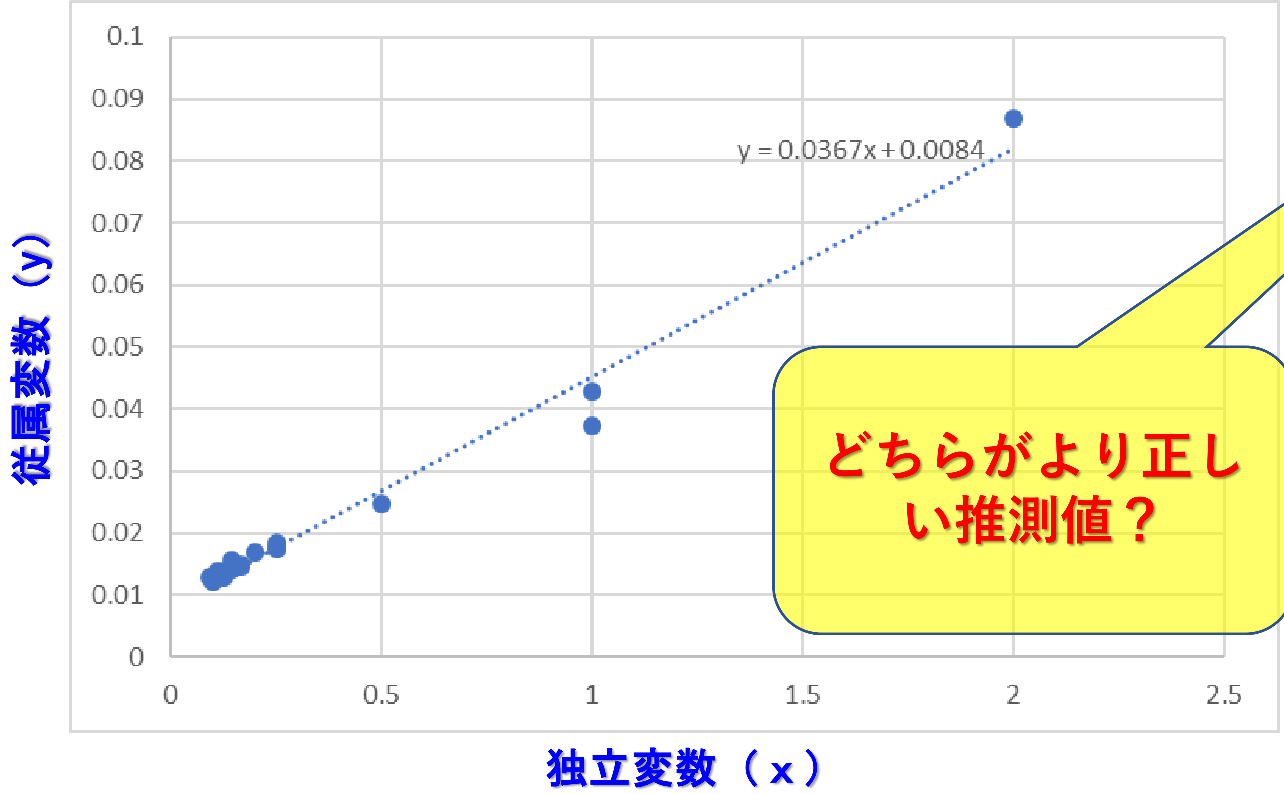


$$V = \frac{V_m x}{K_m + x}$$

両対数プロットからの直線回帰

K_m = 4.3

V_m = 120

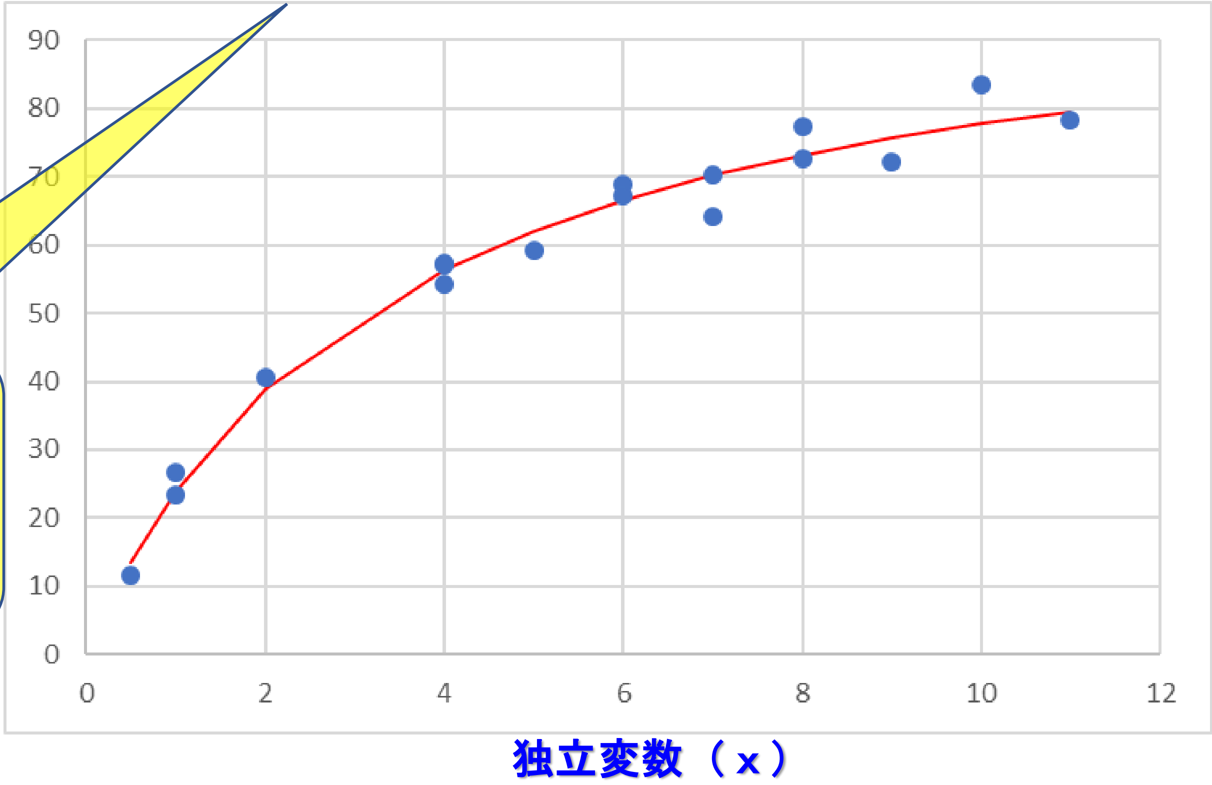


$$y = \frac{1}{V} = \frac{1}{V_m} + \frac{K_m}{V_m} \times \frac{1}{[S]}$$
$$= \frac{1}{V_m} + \frac{K_m}{V_m} \times x$$

曲線への最小二乗法による回帰

K_m = 3.3

V_m = 100

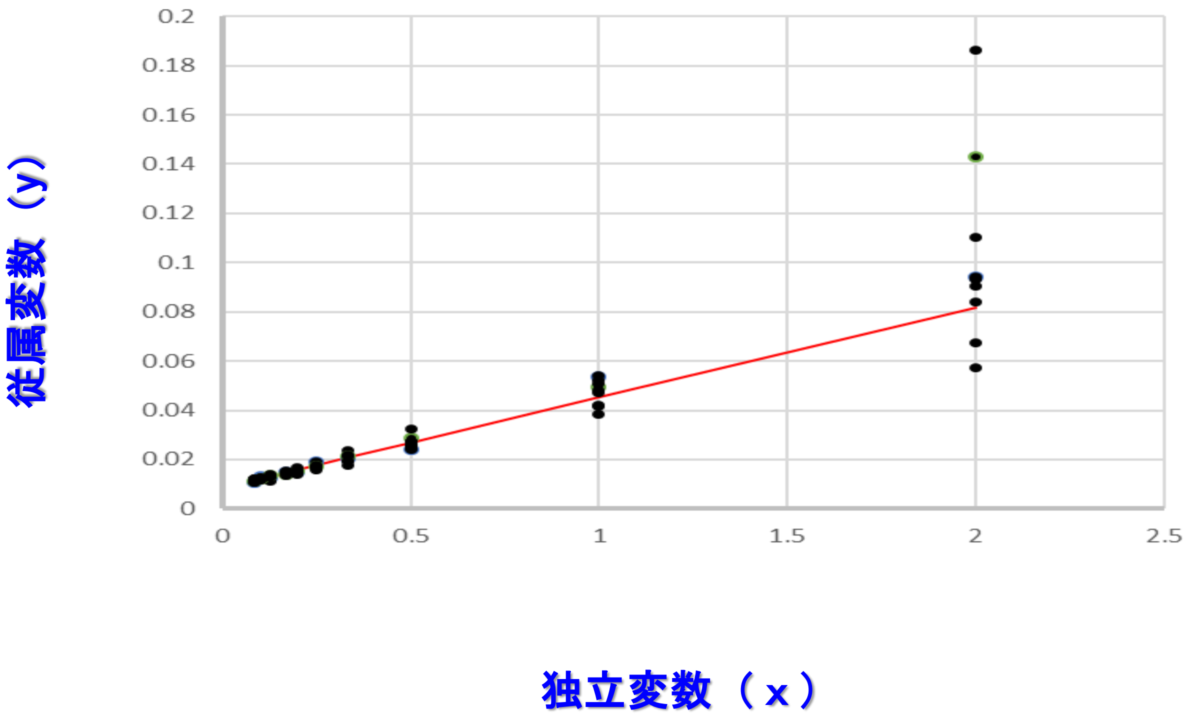


$$V = \frac{V_m x}{K_m + x}$$

どちらがより正しい推測値？

両対数プロットからの直線回帰

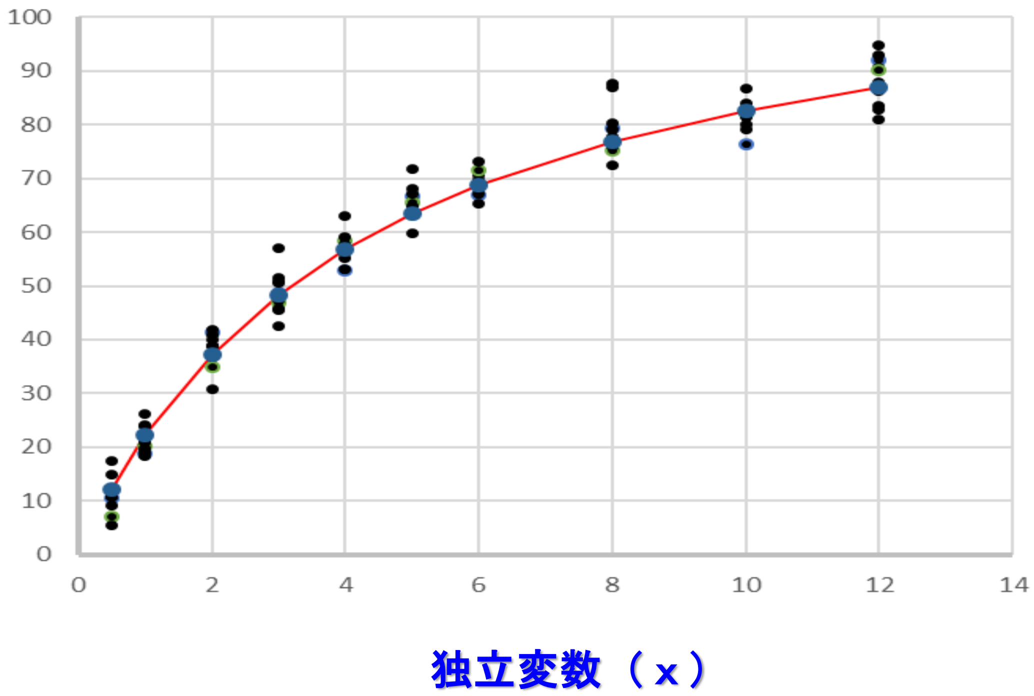
低い基質濃度での測定誤差が大きく影響



$$y = \frac{1}{V} = \frac{1}{V_m} + \frac{K_m}{V_m} \times \frac{1}{[S]}$$
$$= \frac{1}{V_m} + \frac{K_m}{V_m} \times x$$

曲線への最小二乗法による回帰

全基質濃度での測定値を均等に扱う



$$V = \frac{V_m x}{K_m + x}$$