信学会2025総合大会
（AP研シニアセッション）
28 March, 2025

観測データを読む
～間違った判断をしないための～

唐沢　好男

むかしむかし

ばらつきの大きい伝搬データに、
ベテランがエイヤっと線を引く。
そこに名人芸を見た

でも、今の時代、これではいけない。

間違った結論を出さないために、
より良いモデルを見出すために
統計的手法を身につけることが大事

以下、私の実践例

【目的】

地球温暖化時代、日本の雨の降り方は変わってきたか？

気象庁のデータ（ホームページに公開されている膨大なデータ）を調べてみよう

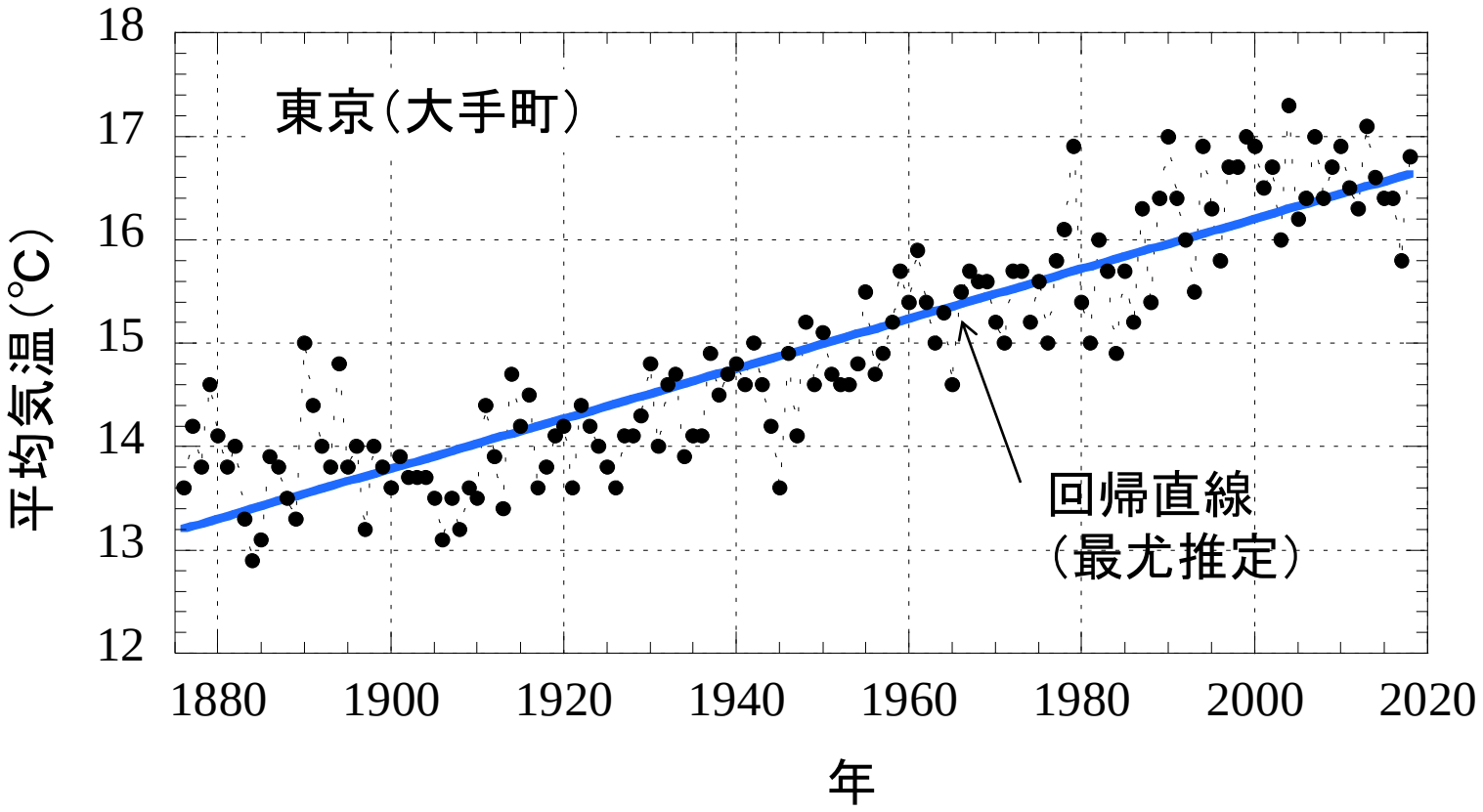
【統計的手法】

データの傾向を掴む：　最尤推定（最小二乗誤差推定）

間違った結論を出さない（結論を出すには慎重に）：　区間推定

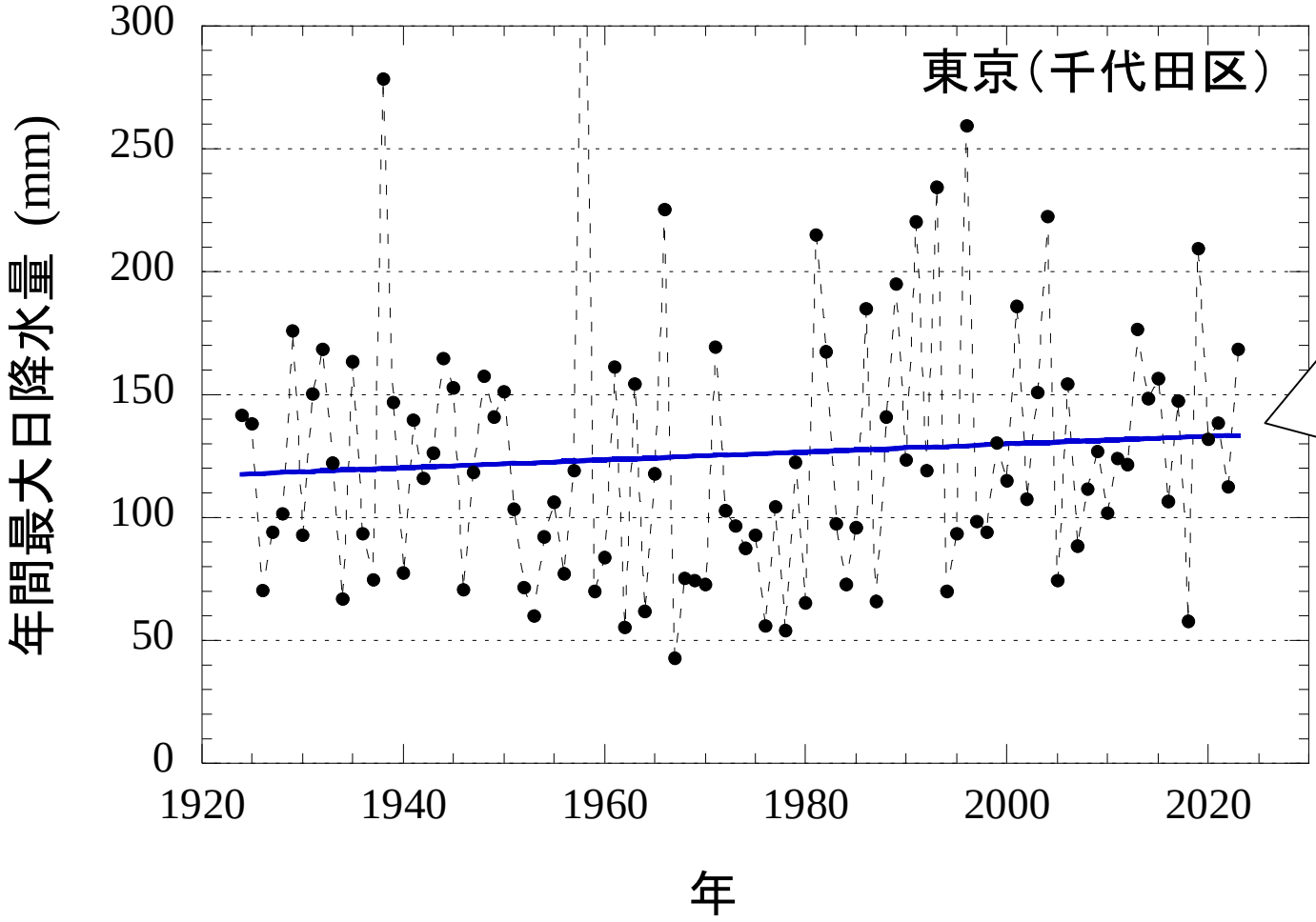
よりよいモデルを見出すために：　赤池情報量基準（AIC）

地球温暖化問題：それは確実にある



東京（大手町にある東京管区気象台エリア）は都市化率（92.9%）、気温の上昇率ともに、日本の都市の中で最大であり、ヒートアイランド現象が最も強く現れているエリア　　約 3℃/100年

雨の降り方はどうか
（東京における年間最大日降水量の100年間の推移）

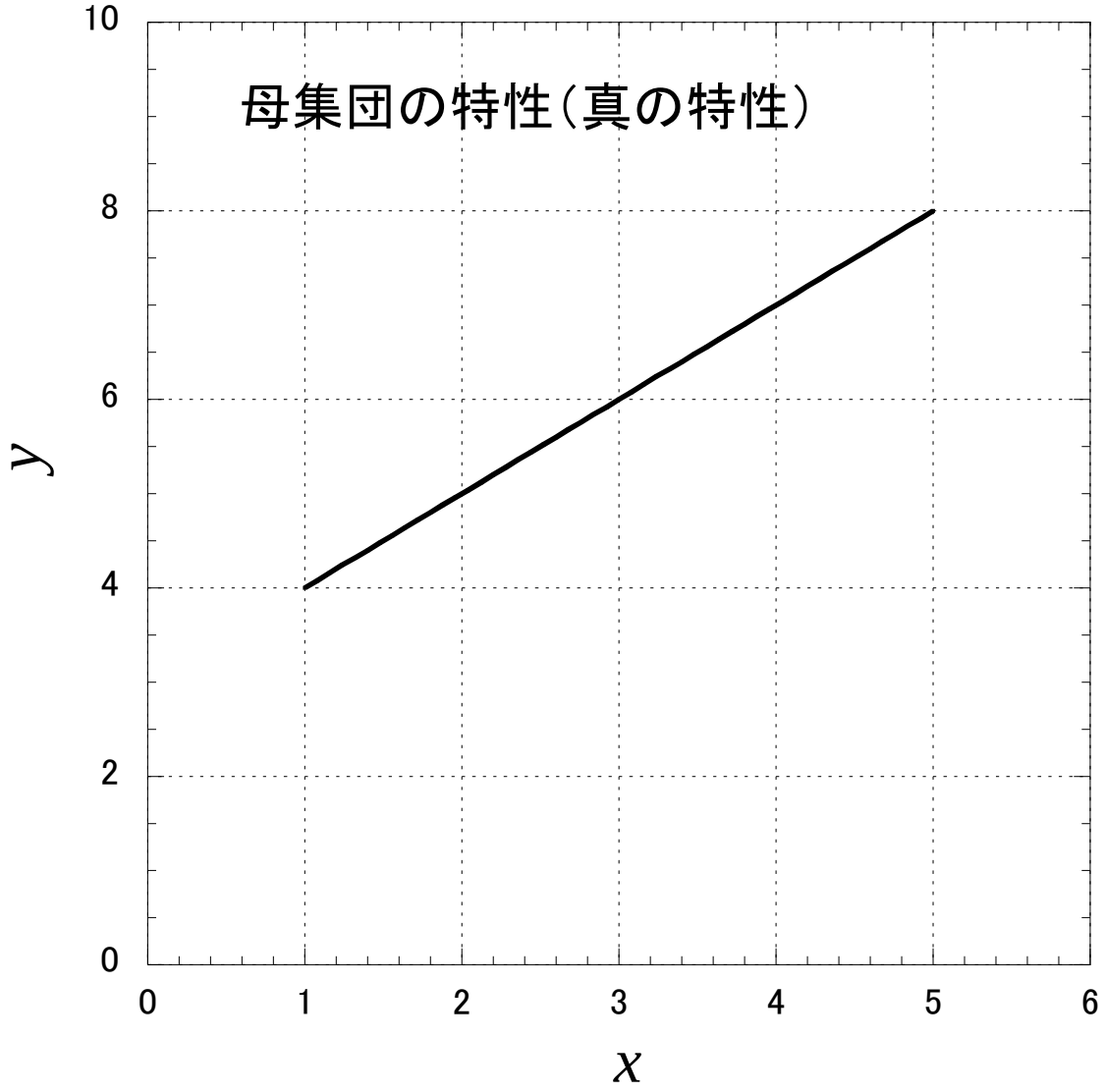


回帰直線
（最尤推定）

増加傾向に
見えるが、
本当にそうか？

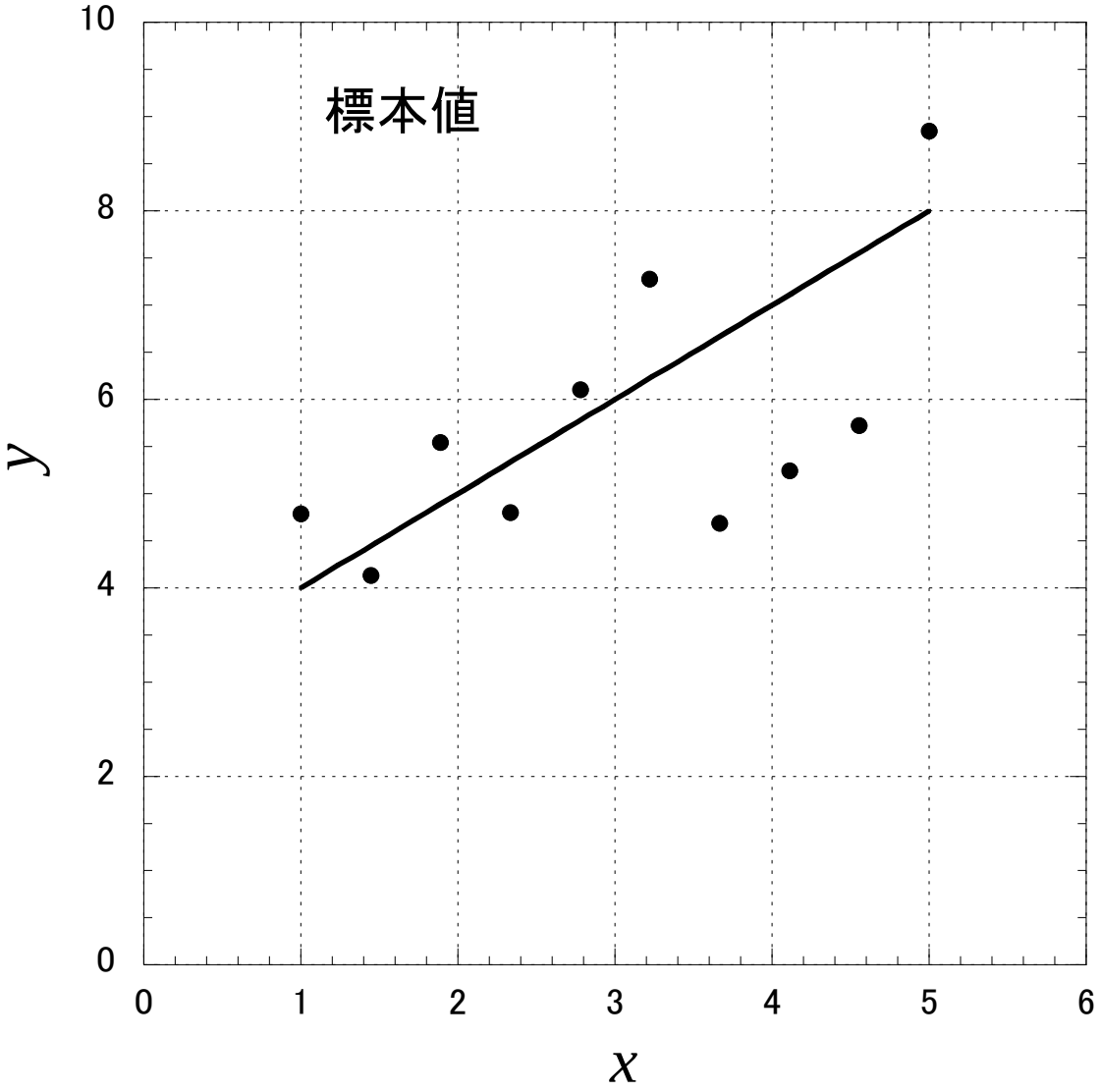
区間推定
をしてみよう

解析の道具： 直線回帰と信頼区間



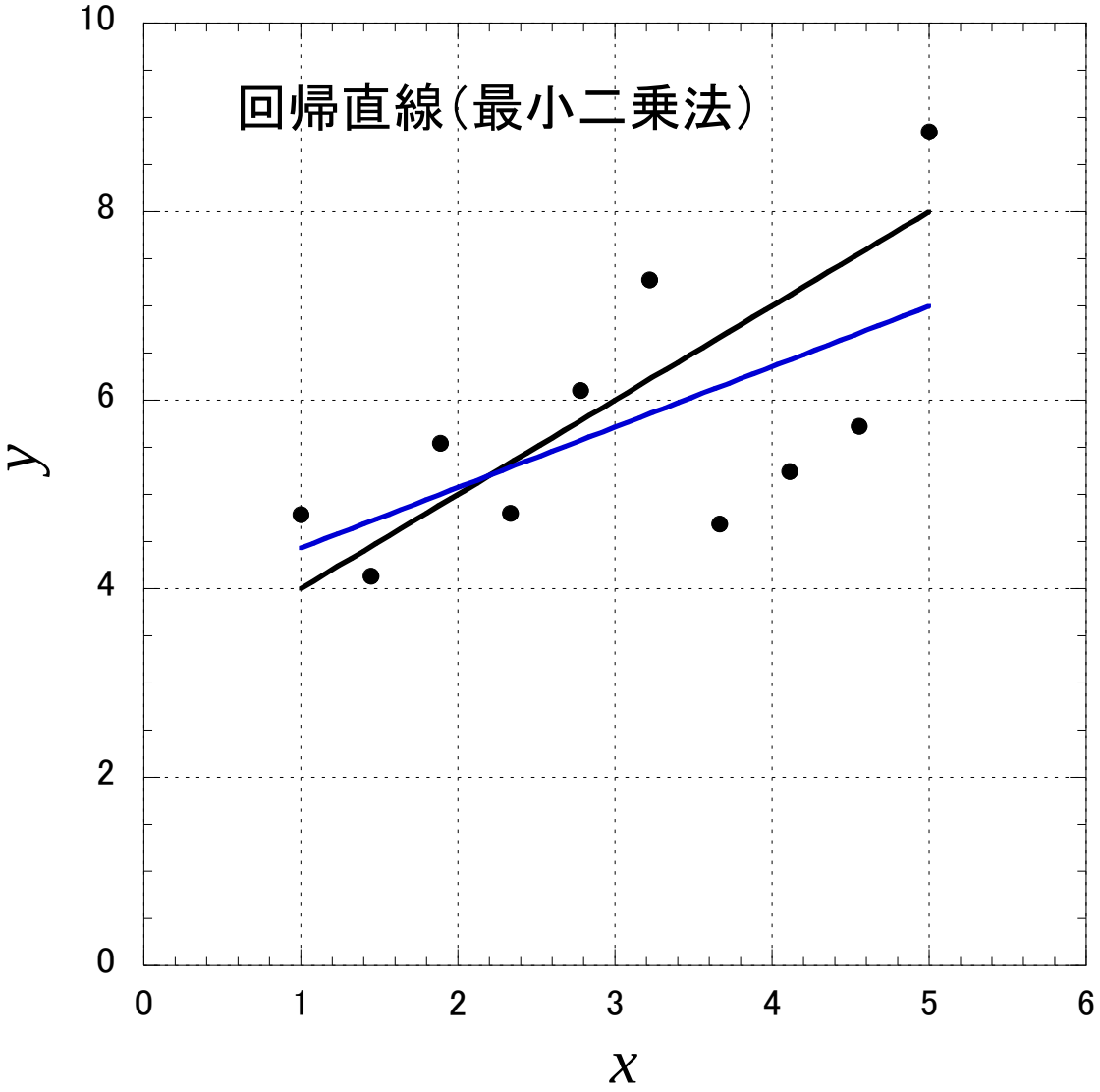
真の特性を
我々は知らない

解析の道具： 直線回帰と信頼区間



知りえるのは
標本値のみ

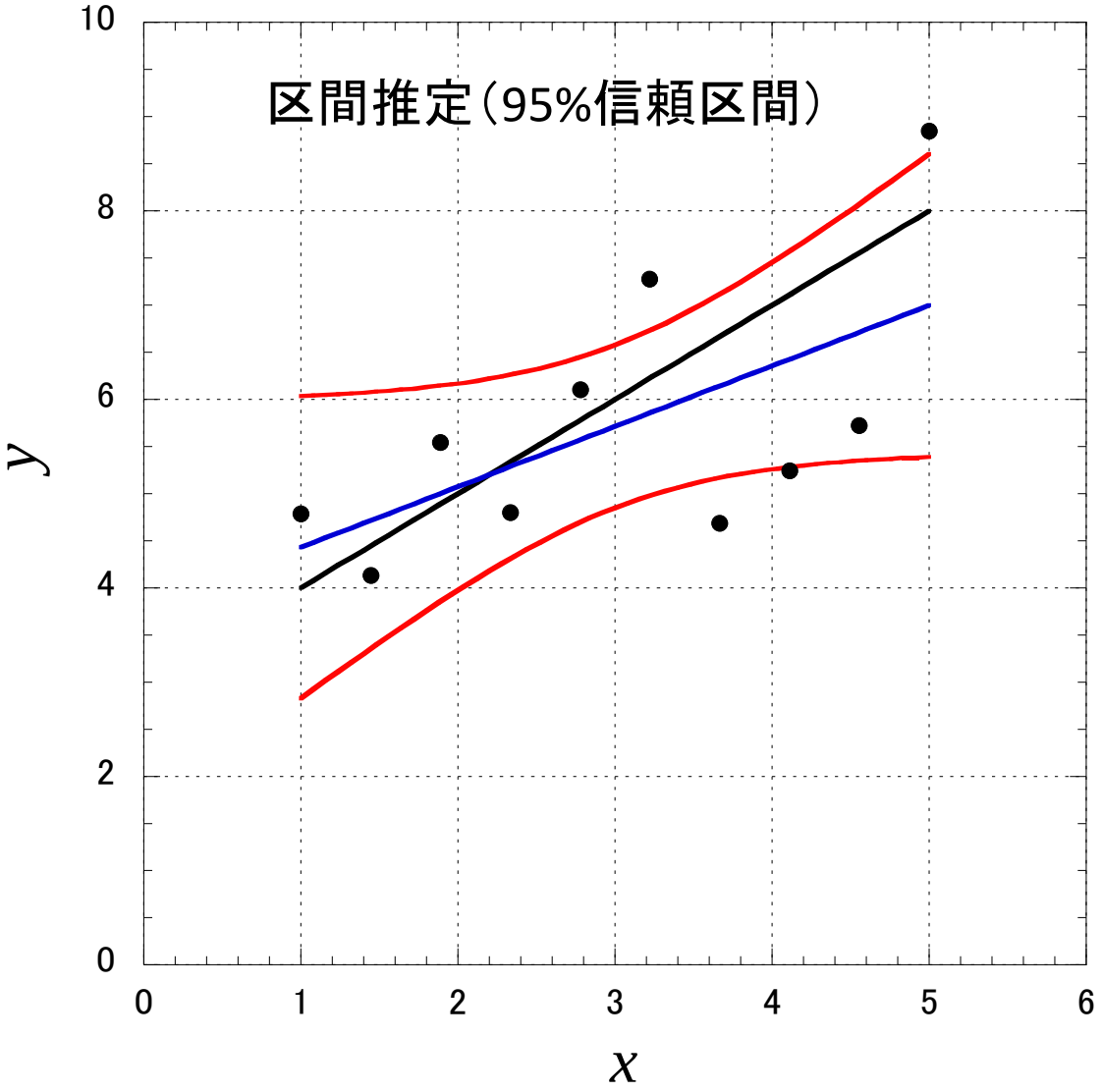
解析の道具： 直線回帰と信頼区間



標本値から
真の特性を推定する

最小二乗法は最尤推定である。ただし、それがどの程度正しいかは分からない

解析の道具： 直線回帰と信頼区間



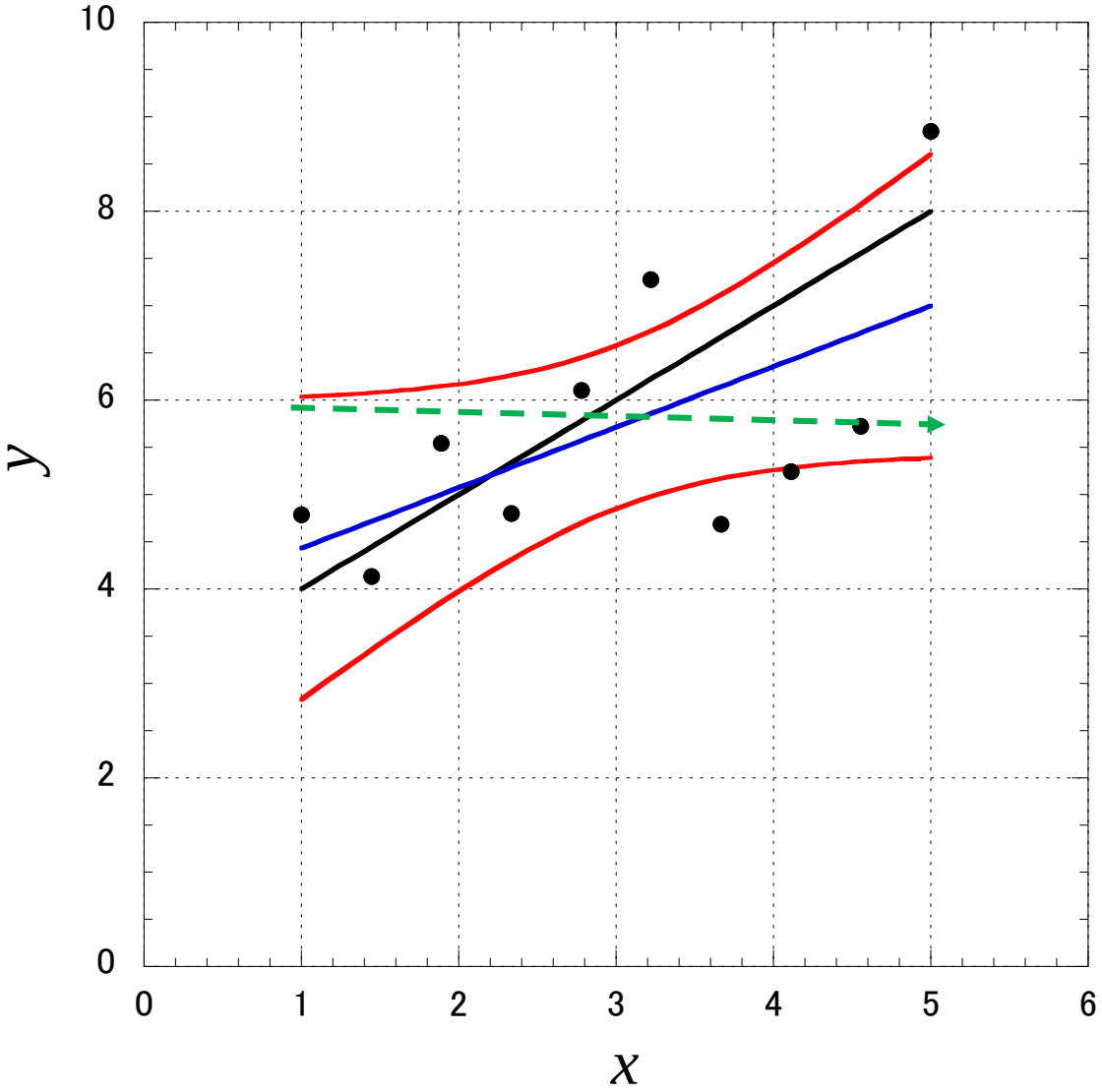
区間推定：
真の特性が存在するであろう確率を定めて、区間で表す

確率値は
90%, 95%, 99%
等

以下では95%基準を採用

信頼区間の計算法は教科書で学んでほしい

解析の道具： 直線回帰と信頼区間



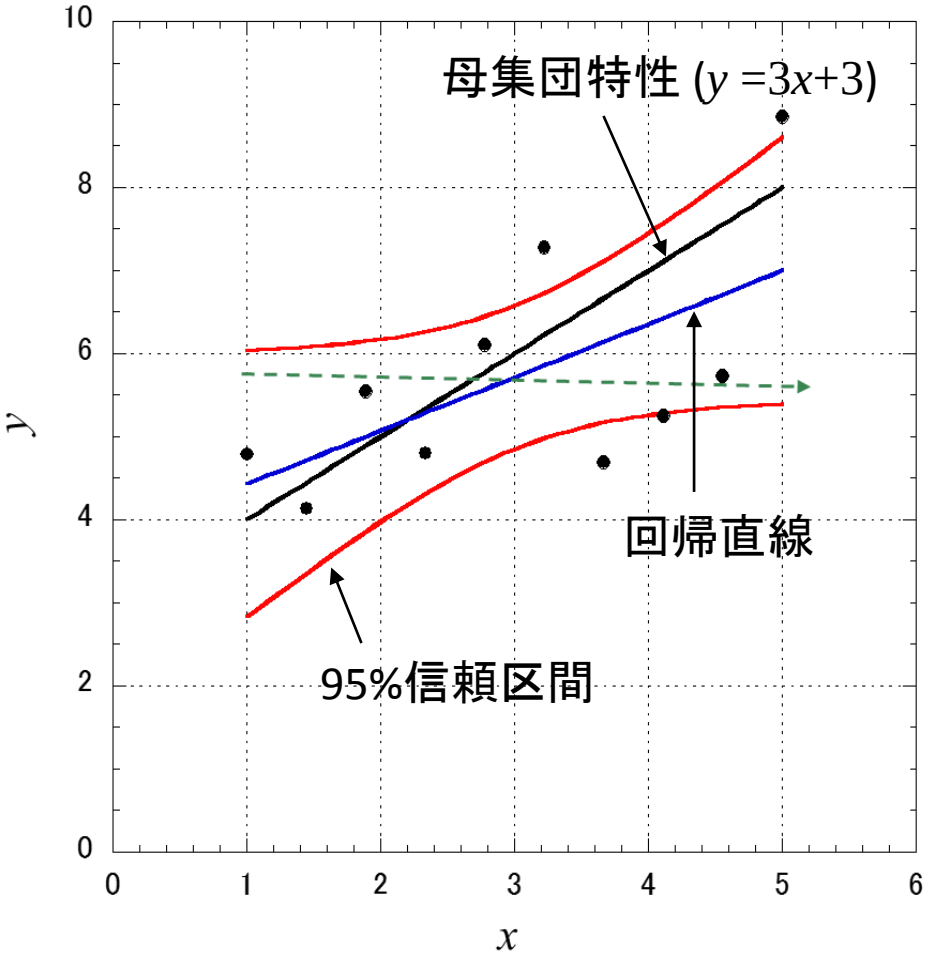
95%信頼区間の中にある直線（緑色の線）は、5%以上の確率があり、棄却できない

この標本値だけでは、xに対して、「増加傾向が有意である」とは言えない

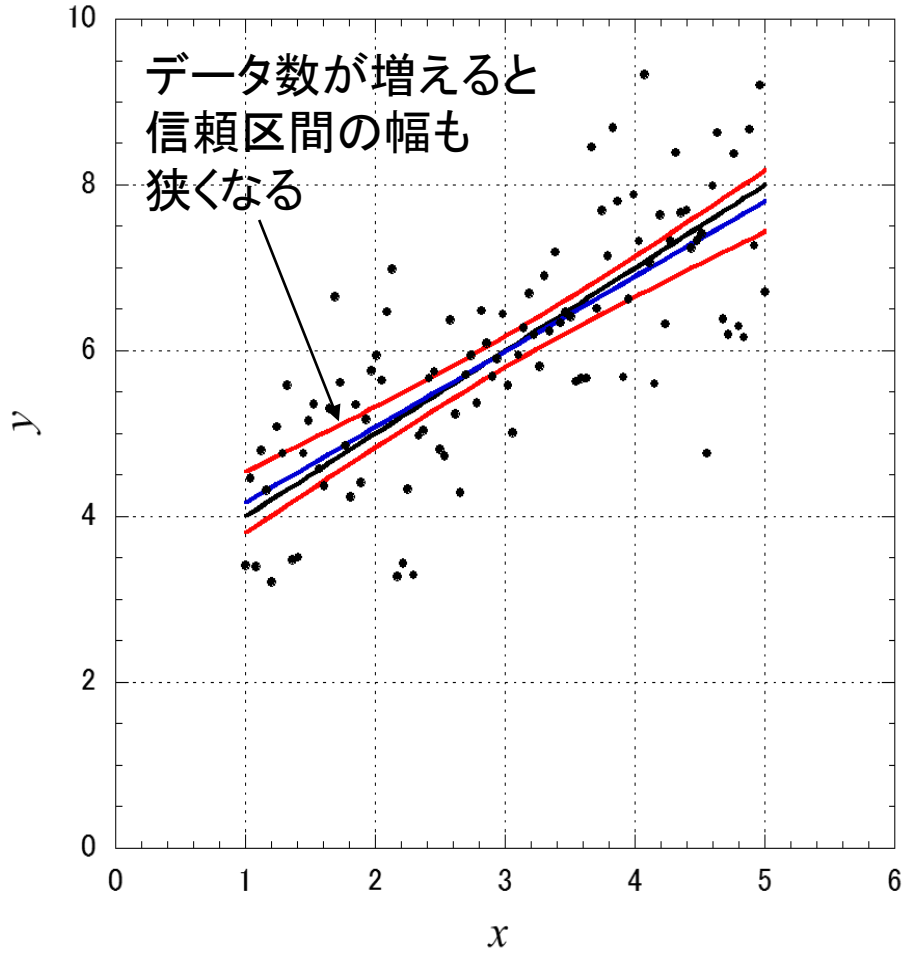
間違った判断をしないための慎重さ重視の推定

解析の道具： 直線回帰と信頼区間

データ数 10

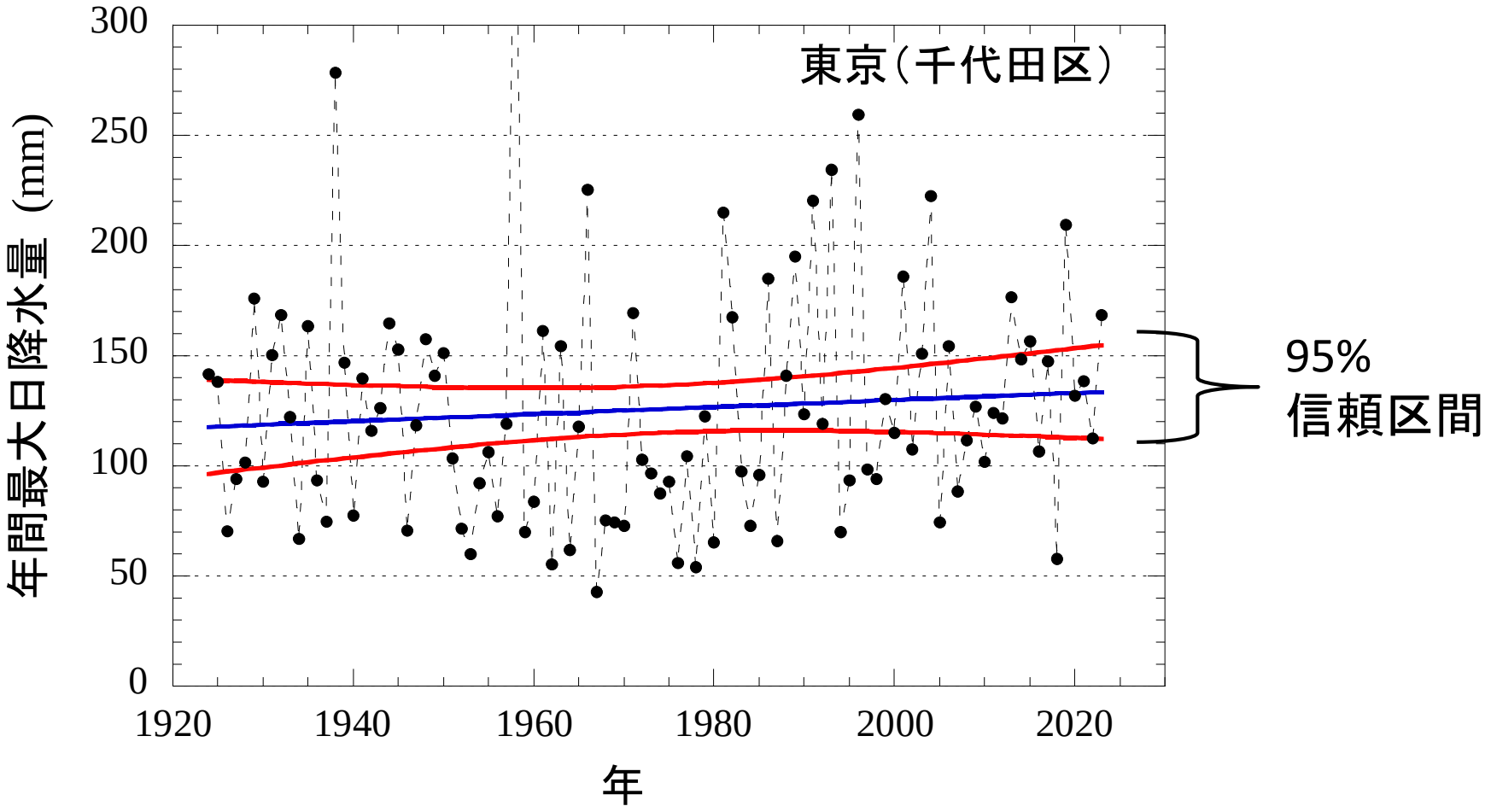


データ数 100



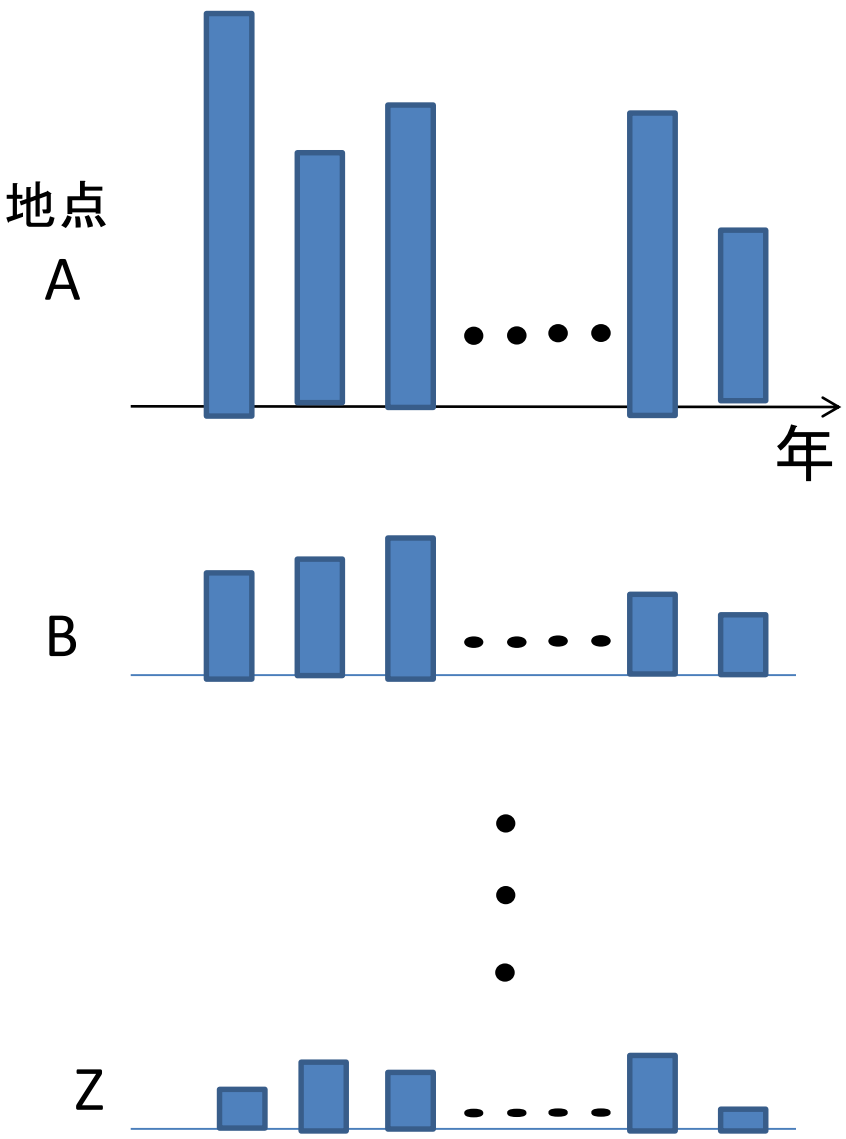
【信頼区間はデータのばらつき幅とはまったく違う!!】

東京における年間最大日降水量の100年間の推移
区間推定（95%信頼区間）をしてみよう



回帰直線（青線）はやや増加の傾向を示しているが、
このデータからは、「増加している」と言う結論は出せない

地点毎に最大降水量の大きさの
平均値が異なる



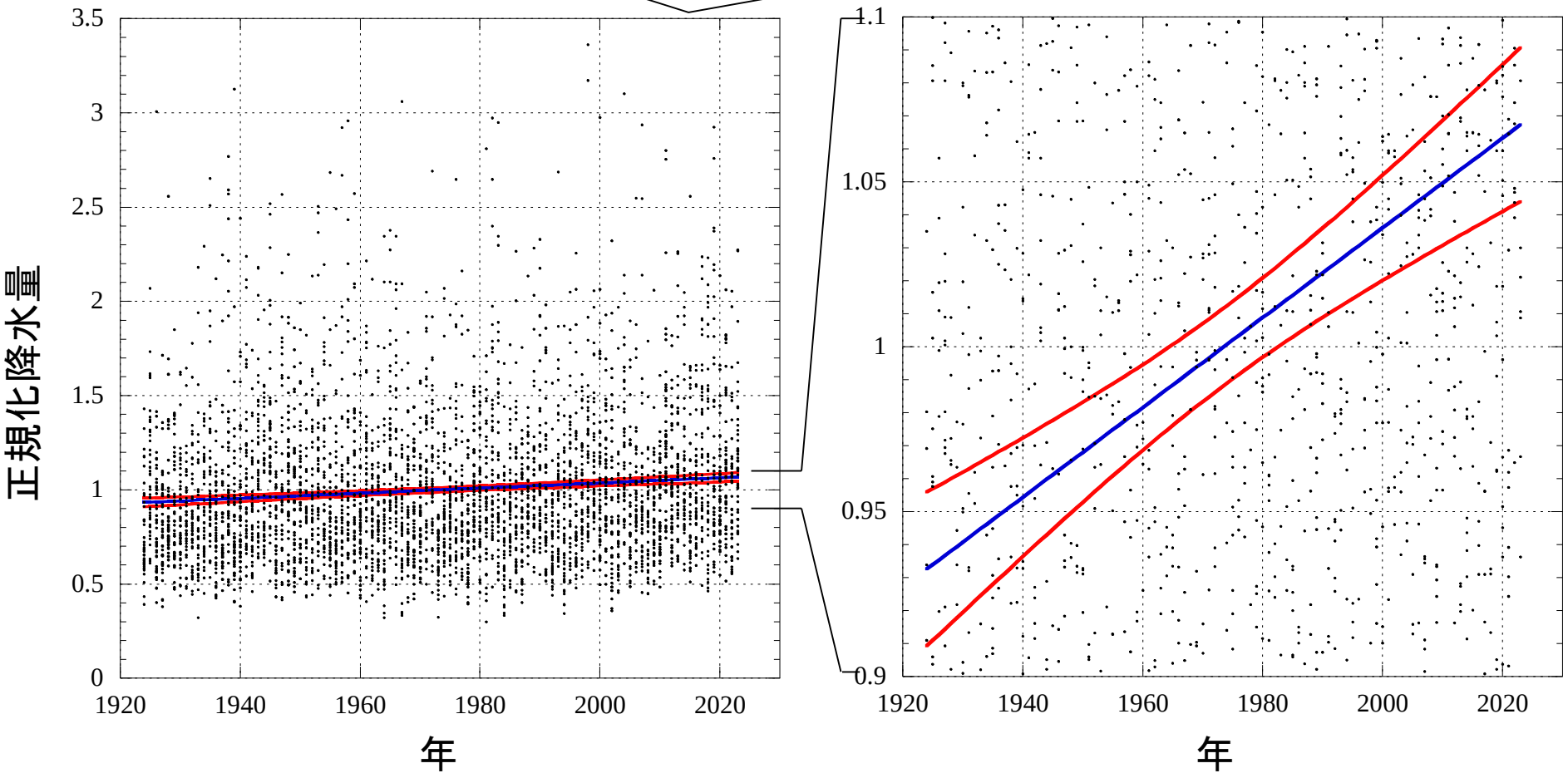
大きさの平均値を同じにする
（長期変化を共通のデータとして扱える）

日本全国をひとまとめにして
データ数を増やす

規格化
（平均値
で割る）

日降水量

日本全国45地点の100年間分のデータ（データ点数：4500）



回帰直線（最尤推定）： 100年間で14%の増加

区間推定（95%信頼区間）：9%～20%の増加の範囲 → 増加傾向が有意

日降水量年間最大値の長期変化傾向を調べたい

区間推定によって、ようやく見えてきた増加傾向
（95%信頼区間の推定で、少なくとも100年間で10%の増加）

↓

長期の変化をより詳しく見てみたい
多項式近似をしよう

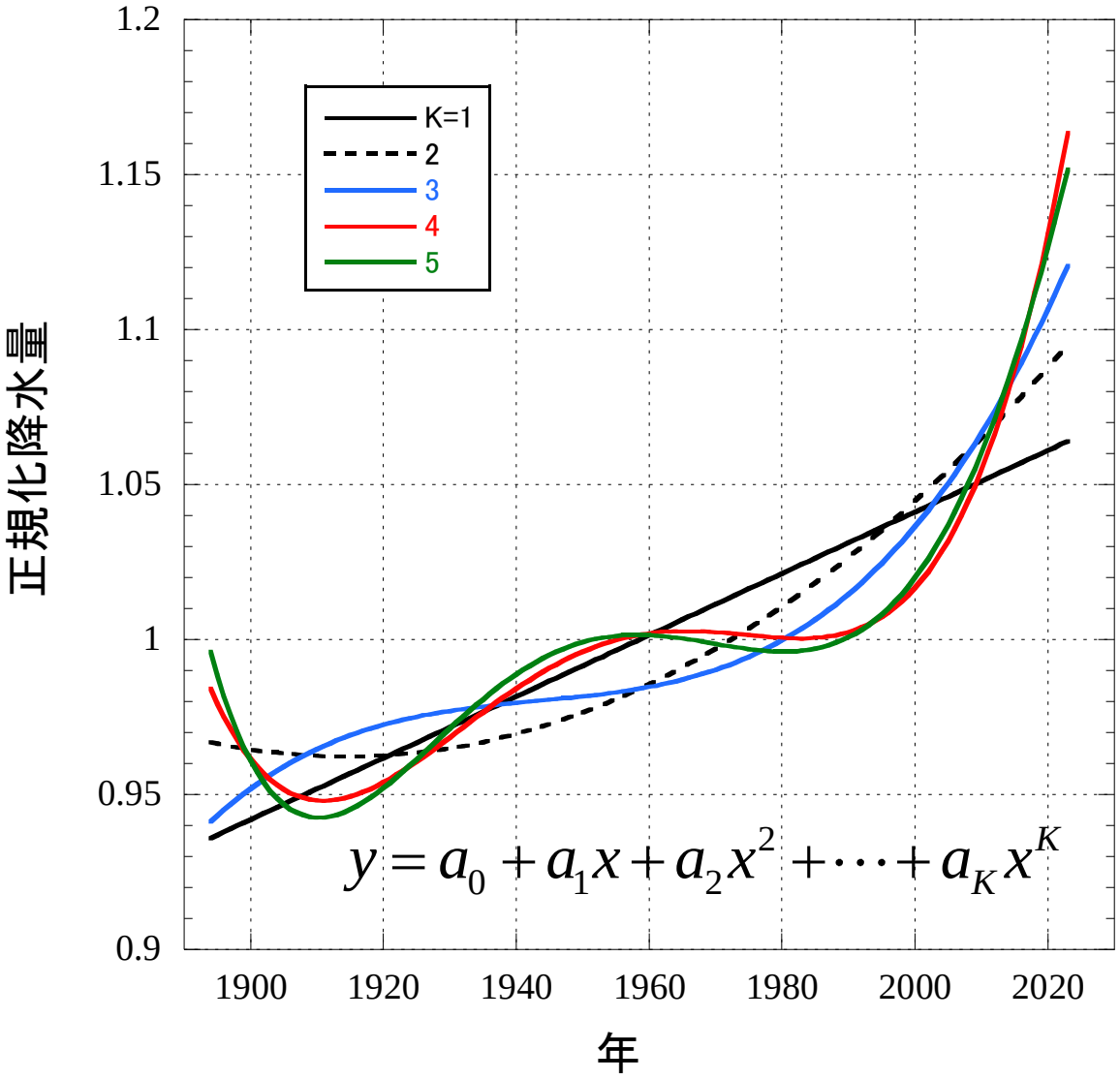
↓

では、どの次数を選ぶのがよいのか?

↓

AIC（赤池情報量基準）を使おう

正規化日降水量の130年間の変化の多項式近似による最尤推定



次数を上げるほど、詳細な変化が把握できる

⇩

次数を上げすぎると、標本データの誤差要因に支配されて、他の独立なデータに合わなくなる

⇩

最適な次数をどのように定めるべきか

⇩

AIC（赤池情報量規準）で決めよう

最適なモデルを選ぶ指標：赤池情報量基準（AIC）

標本データから、母集団を推定する良いモデルを選ぶ判断基準

赤池弘次博士（（元）統計数理研究所：1971年考案）

良いモデルとは
1） モデル（推定式）と標本値の差が小さい（データによく合う）
2） モデルのパラメータの数が少ない

AIC= (-2)（最大対数尤度）＋2（パラメータ数）

AICの値が小さいほど良いモデル

最大対数尤度とパラメータ数が同じ次元になっているのが奇跡
この基準は今でも現役で、広く使われている
理論は難しいが、確率・統計知識の宝庫

AICによる多項式近似の評価

AICにおけるモデルの選択基準

$$AIC = -2\times(\text{最大対数尤度}) + 2\times(\text{パラメータの数})$$

の式において、最も値が小さくなるモデルを選ぶ
（尤度が同程度ならパラメータ数が少ないほうが良いモデル）

K次多項式近似のモデル化（誤差が正規分布する）

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_K x_i^K + b_i \quad \left(f_b = N(0,\sigma^2)\right) \quad （\text{パラメータ数}: K+2）$$

正規分布の対数尤度関数

$$l(\boldsymbol{\theta}) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \sum_{k=0}^{K} a_k x_i^k\right)^2 \qquad \boldsymbol{\theta} = \left(a_0, a_1, a_2, \cdots, a_K, \sigma^2\right)$$

係数と誤差分散の最尤推定 $\quad \dfrac{\partial}{\partial a_i}\sum_{i=1}^{n}\left(y_i - \sum_{k=0}^{K} a_k x_i^k\right)^2 = 0 \quad \Rightarrow \ \hat{a}_0 \sim \hat{a}_K$

$$\frac{\partial l\left(\boldsymbol{\theta}(\hat{a}_0, \cdots, \hat{a}_K, \sigma^2)\right)}{\partial \sigma^2} = 0 \ \Rightarrow \ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{k=0}^{K}\hat{a}_k x_i^k\right)^2$$

最大対数尤度 $\quad l(\hat{\boldsymbol{\theta}}) = -\dfrac{n}{2}\log\left(2\pi\hat{\sigma}^2\right) - \dfrac{n}{2}$

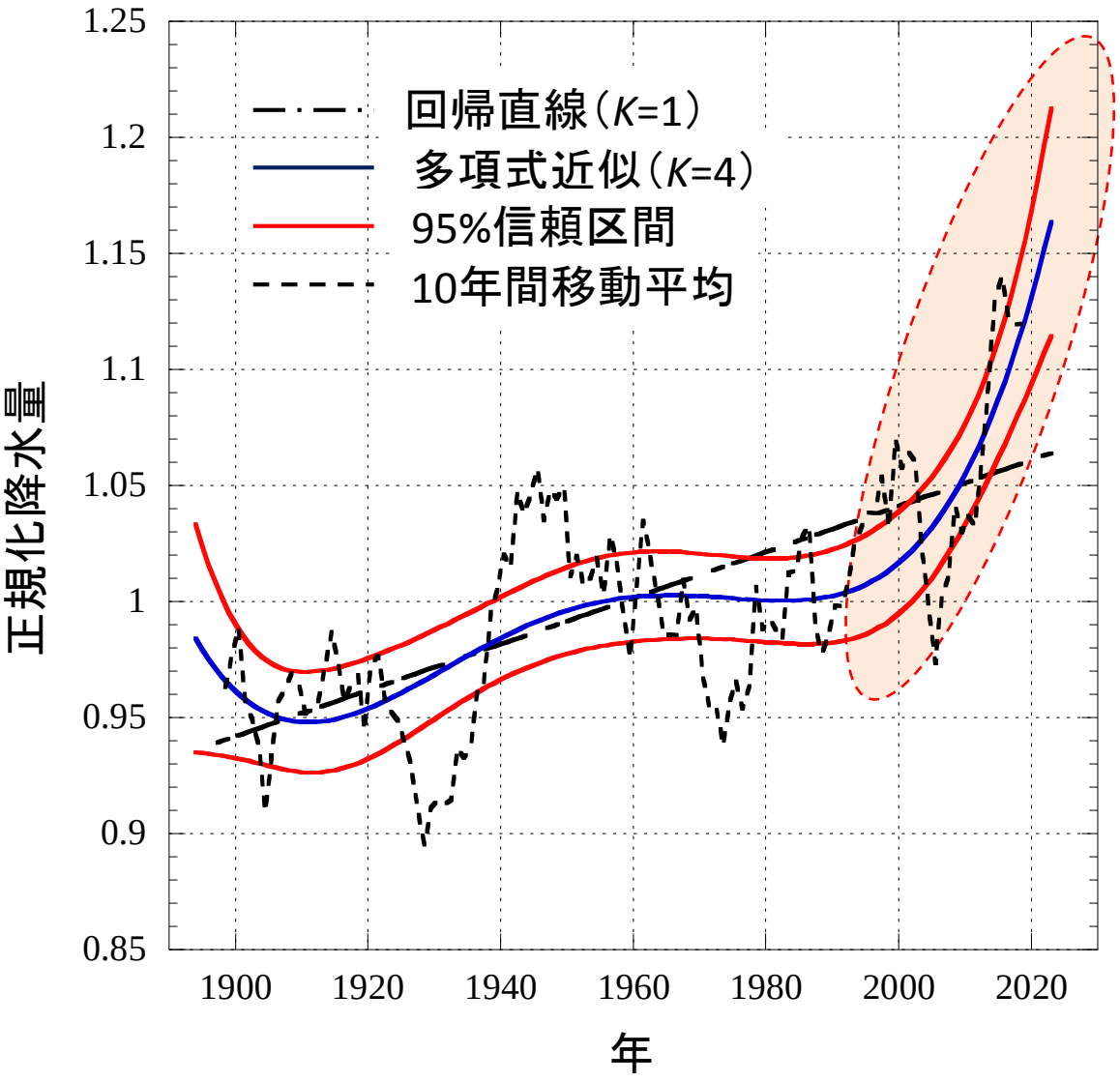
多項式近似における最大対数尤度とAIC評価値

値が最小のものを選ぶ

Rainfall	K	$\hat{\sigma}^2$	$l(\hat{\theta}^2)$	AIC	ΔAIC
1-day (130年)	1	0.16089	-2956.79	5919.57	13.79
	2	0.16069	-2953.13	5914.26	8.472
	3	0.16059	-2951.24	5912.49	6.704
	4	0.16035	-2946.89	5905.78	0
	5	0.16033	-2946.60	5907.19	1.409

選ばれた次数

日降水量年間最大値の長期変化傾向
（AICは4次式近似を最適と判断）



1990年ころから
増加傾向

他のデータも含めて
現在、詳細、調査中

まとめ

いろいろの統計手法を身につけておくと、
研究や開発を進める上で役に立つ武器になる

持ち駒を多くして、どのような手法がどのようなときに役立つかを
知っておくとよい

手法の詳細（具体的な計算法）は必要になったとき
しっかり学べばよい

電波伝搬解析に現れる確率分布と統計的手法をまとめている

http://www.radio3.ee.uec.ac.jp/ronbun/TR-YK-078_Probability_Distributions.pdf