マーケティング・リサーチ特論
～統計的モデリングとは？～

2026年度1学期:　　金曜3限

担当教員:　　　　　石垣　司

統計的モデリング

確率モデル

- 確率現象を数学(確率論)で記述したモデル
  - 標本空間, 確率変数を含む確率分布などを指す場合が多い

統計モデル

- 確率的構造を伴う対象や現象の数理モデル

統計的モデリング

- より良い統計モデルを作る・探求する行為

マーケティング・リサーチにおける統計的モデリング

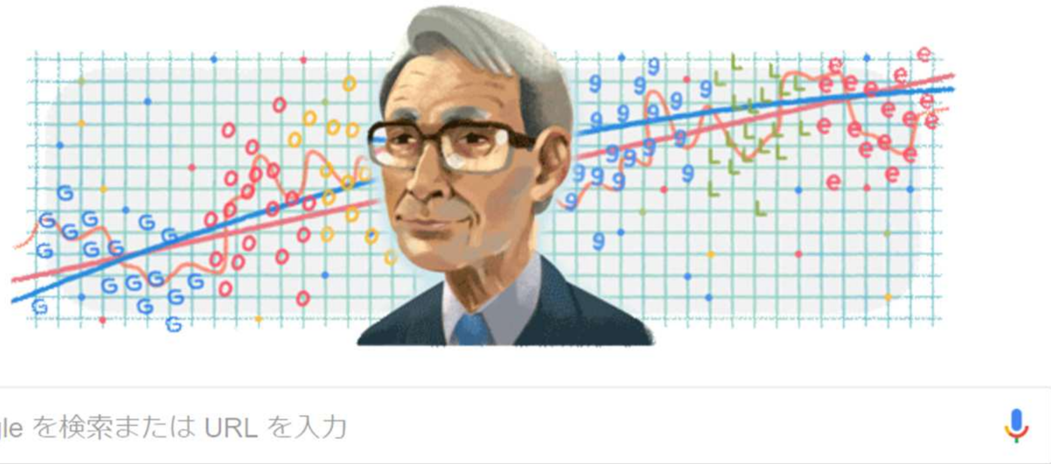
- 購買行動や消費者心理の分析において, 直接観測することができない潜在変数・構成概念の推定に有効
  - アンケート結果の単純集計からの脱却

統計科学とモデル #1

赤池弘次『時系列解析の心構え』 *時系列解析の実際*（赤池、北川編）朝倉書店 1995

統計的方法の本質は，データを用いて必要な情報を創り出すことにある。（中略） 統計科学は，このような情報の創出の過程に関する人間の知的活動の科学と定義される

モデルは仮説の表現であり，仮説の提案こそが基本的な知的活動である。これによって，データに意味が発生し，情報が創り出されるのである



2017年11月5日のGoogleロゴ
https://www.google.com/doodles/hirotugu-akaikes-90th-birthday?hl=ja

統計科学とモデル #2

赤池弘次『時系列解析の心構え』 *時系列解析の実際*（赤池、北川編）朝倉書店 1995

確率の説明は様々になされるが，**その本質は我々の持つ期待の構造の形式的な表現**である。

（従来の統計的推定論や検定論に対し）これらの議論の根底に有るのは，いわゆる"真の構造"あるいは"真理"の考えである。（ところが，特殊な場合を除き）期待の構成の仕方は我々の持つ知識や経験の使い方に大きく依存する。従って，**唯一無二の真の構造のようなものは存在しない**。

期待は（中略）環境あるいは対象の特性を部分的に表現する形のモデルを用いて表現される。したがって，**我々はより良いモデルの探求を通じて，常に未知の状態にある究極的な真理あるいは真の構造に迫るのである**

（　）の中は担当教員による補足

より良いモデルの探究

モデルは常に完璧ではない

– “ Essentially, all models are wrong, but some are useful.”
(G.E.P. Box, Box-Jenkins methodology)

より良いモデルの探求

– より本質的な対象のメカニズムを説明したい
– より良い予測・制御のために, より本質的な情報を得たい

⇒人間の知的活動としての統計的モデリング

統計的モデリングによる情報の創出には？

– 問題の対象に対する深い理解
– 問題に適した<u>統計モデル</u>
– 統計モデルに適したパラメータの<u>推定原理</u>と<u>推定手段</u>

モデル・推定原理・推定手段

データを用いた情報の創出には3つの調和が必須

モデル

対象の特性の部分的な表現。（本授業では主に）確率的構造の数学的表現

-本授業で扱うモデル-

線形回帰モデル,
因子分析モデル,
多項ロジットモデル,
階層回帰モデル, など

推定原理

未知のパラメータを含んでいるモデルとデータを整合させるための原理

-本授業で扱う原理-

① ズレの最小化
（最小2乗推定, 一般化モーメント法, LASSOなど）
② 尤度の最大化
③ 事後分布の推定

推定手段

モデルとデータを整合させるための関数やアルゴリズムなどの手段

-本授業で扱う手段-

推定量, ニュートン法,
EMアルゴリズム,
マルコフ連鎖モンテカルロ法, など

※3つは混同しやすいので注意。特に原理と手段に関しては“推定法”などとまとめて書かれている書籍等もある。ここでは混同を防ぐため, 推定原理と推定手段という言葉で区別する。

例えば、“階層ベイズモデル”は階層的モデルをベイズ推定の原理で扱うということ。モデルと原理を両方含んでいる言葉。経済学やマーケティング・リサーチで扱う階層ベイズモデルは回帰モデルや離散選択モデルがほとんどだが、他の分野ではそうとは限らないので注意

これ以降の変数の記述法 #1

スカラ：小文字, **ベクトル：小文字&太文字**, **行列：大文字**

- **例：** $x = 1,\ \boldsymbol{x} = [1,2,1,\cdots,3]^T,\ X = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$

目的変数： y

説明変数ベクトル： $\boldsymbol{x} = [1, x_1, x_2, \cdots, x_P]^T$

データ： $D = \{y_i, \boldsymbol{x}_i\}\ (i = 1, \cdots, N)$

- $\boldsymbol{x}_i = [1 \quad x_{i1} \quad \cdots \quad x_{iP}]^T$ **（行列** X **の** i **番目の行）**
- $X = [\mathbf{1} \quad \boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_P]^T = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1P} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \ddots & x_{iP} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NP} \end{bmatrix}$

回帰係数：係数 $b_0, b_1, \cdots, b_P$ **（パラメータ）**

- $\boldsymbol{b} = [b_0 \quad b_1 \quad \cdots \quad b_P]^T,\ b_0$ **は切片項の係数**

誤差項： e_i

これ以降の変数の記述法 #2

線形重回帰モデル

$$\begin{cases} y_1 = b_0 + b_1 x_{11} +, \cdots, + b_P x_{1P} + e_1 \\ \vdots \\ y_i = b_0 + b_1 x_{i1} +, \cdots, + b_P x_{iP} + e_i \\ \vdots \\ y_N = b_0 + b_1 x_{N1} +, \cdots, + b_P x_{NP} + e_N \end{cases}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1P} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \ddots & x_{iP} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NP} \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_P \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_i \\ \vdots \\ e_N \end{bmatrix}$$

線形回帰モデルの行列表現

$$\boldsymbol{y} = X\boldsymbol{b} + \boldsymbol{e}$$

モデルは我々の持つ期待の数理的表現

線形回帰モデルを例に

- 例：顧客満足度アンケート調査

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + e_i, e_i \sim N(0, \sigma^2)\ (i = 1, \cdots, N)$$

y: 顧客満足度（0~100点）, x_1: 価格, x_2: ブランド知名度

- このモデルが表現する消費者行動に関する仮説（期待）
  1. 価格は顧客満足度に影響を与える
  2. ブランド知名度は顧客満足度に影響を与える
  3. 説明変数の変化は目的変数の変化に線形的に影響を与える

- 不確実性の表現（確率的構造）

各個人 i に対して, 仮説のみでは表現できない顧客満足度のブレは確率的構造(e_i)に従う

より良いモデルへの改良

1. **線形回帰モデルの枠組みで改良**
   – **対象に対する知識や経験を投入**

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + e_i, e_i \sim N(0, \sigma^2)$$

y: **顧客満足度**, x_1: **価格**, x_2: **ブランド知名度**, x_3: **接客時間**

「接客時間は顧客満足度に影響を与える」という期待を追加

2. **モデルの形や確率的構造の改良**

$$y_i = f(x_{1i}, x_{2i}, x_{3i}) + e_i, e_i \sim P(\cdot)$$

   – **説明変数と回帰係数の線形結合以外の表現**
   – **誤差項の確率的構造が非正規分布**

モデルの形自体の改良が必要な例 #1

目的変数が2値の離散変数 {0,1} の回帰分析

– 例：消費者による商品の購買/非購買の選択モデル

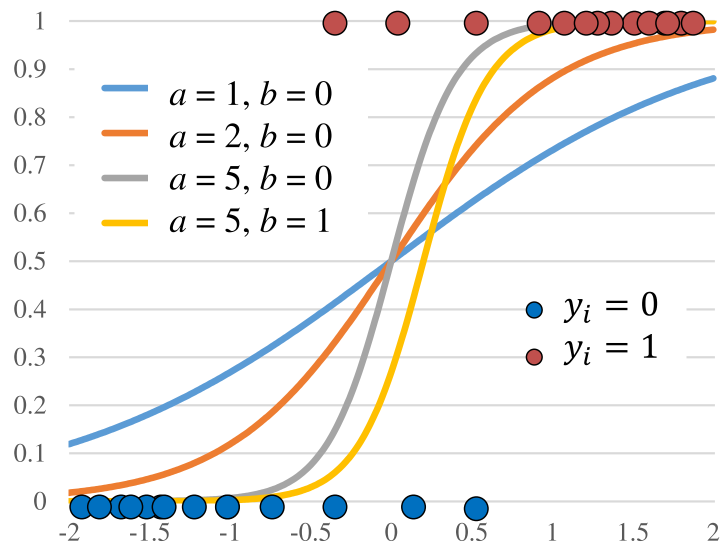
ロジスティック回帰分析

– 関数形：$f(x) = \dfrac{1}{1+e^{-bx}}$

「推定」とはデータをうまく表現できる
S字カーブの形状を決めること

【抽出できる情報の例】
商品Aが250円の時の購買確率70%（30%非購買）
商品Aが300円の時の購買確率20%（80%非購買）

$$f(x) = \Pr(y = 1): 0 \leq \frac{1}{1 + e^{-(a+bx)}} \leq 1$$



線形回帰分析の関数形（回帰直線）では
データにうまくフィットしない
S字型の関数形が必要。

モデルの形自体の改良が必要な例 #2

目的変数が名義尺度の場合の回帰分析

– 例：消費者によるブランド A, B or C の選択モデル

多項ロジットモデル（詳細は本授業の後半の回で）

関数形： $\Pr(Y_i = j) = \dfrac{\exp(\boldsymbol{b}\boldsymbol{x}_{ij})}{\sum_{l=1}^{J} \exp(\boldsymbol{b}\boldsymbol{x}_{il})}$

$\Pr(Y_i = j)$：**消費者 i によるブランド j の選択確率**

【抽出できる情報の例】

ブランドA・B・Cが200円の時のブランドAの選択確率60%（ブランドB:25%,ブランドC:15%）

ブランドAが200円でブランドB・Cが180円の時のブランドAの選択確率30%

モデルの形自体の改良が必要な例 #3

目的変数が名義尺度で主体毎に係数が異なる回帰分析

– **例：消費者毎に選好が異なるブランドA, B, Cの選択モデル**

階層多項ロジットモデル （詳細は本授業の後半の回で）

関数形：
$$\begin{cases} \Pr(Y_i = j) = \dfrac{\exp(\boldsymbol{b}_i \boldsymbol{x}_{ij})}{\sum_{l=1}^{J} \exp(\boldsymbol{b}_i \boldsymbol{x}_{il})} \\ \boldsymbol{b}_i = \Phi \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \Sigma) \end{cases}$$

消費者　過去のデータ

- 価格
- 内容量
- キャンペーン

etc.

ブランドA

- 価格
- 内容量
- キャンペーン

etc.

ブランドB

- 価格
- 内容量
- キャンペーン

etc.

ブランドC

【抽出できる情報の例】

佐藤さんはブランドAが200円でB・Cが180円の時のブランドAの選択確率10%

鈴木さんはブランドAが200円でB・Cが180円の時のブランドAの選択確率90%

「より良い」モデルの探究

赤池弘次『時系列解析の心構え』 *時系列解析の実際*（赤池、北川編）朝倉書店 1995

「したがって, 我々はより良いモデルの探求を通じて, 常に未知の状態にある究極的な真理あるいは真の構造に迫るのである」

より<u>良い</u>: どのような意味で良いのか？

– Keywords: **予測, 汎化性能, 過学習**

<u>より</u>良い: 複数のモデルの中のどれがより良いのか？

– Keywords: **モデル選択**

「より良い」：どのような意味で良いのか？#1

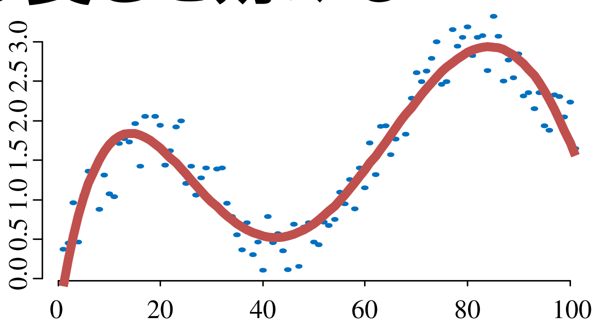
未知データの予測の良さで評価

– 実問題への応用の多くで必要なのは未知データの予測

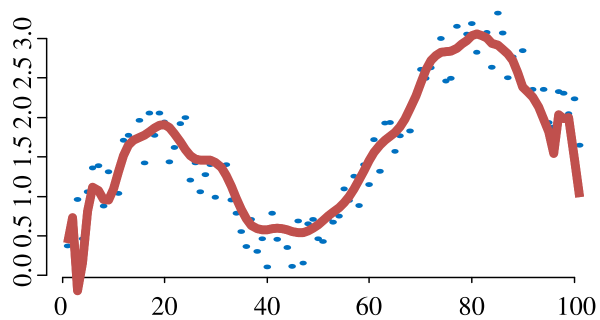
既知データ（取得済みデータ） $\xrightarrow{\text{推定}}$ 統計モデル $\xrightarrow{\text{予測}}$ 未知データ

統計モデル → 既知データ：~~予測~~（多くの場合に意味なし）

– 既知データのみへの過剰適合（過学習）は未知データの予測の良さを妨げる



統計モデルと既知データの残差：中
未知データの予測性能：高



統計モデルと既知データの残差：小
未知データの予測性能：低

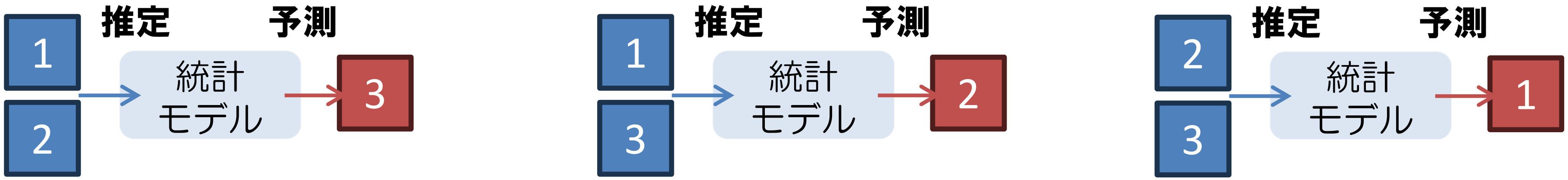
「より良い」:どのような意味で良いのか？#2

統計モデルの未知データへの対応力(汎化性能)の測り方

- クロスバリデーション(交差検証法)

  例：3-fold クロスバリデーション(既知データを3分割)

  この3つの予測性能の平均で統計モデルの良さを評価



- 赤池情報量規準 AIC

  統計モデルの予測分布 $p(x)$ と真の分布 $q(x)$ の近さを KL ダイバージェンスで評価

  KLダイバージェンスの項の一つの平均対数尤度の推定量 AIC が小さいほど $p(x)$ と $q(x)$ が近い(= 良い統計モデル)

AIC = -2 最大対数尤度 + 2 統計モデルの自由パラメータ数

「より良い」:どのモデルがより良いのか？

モデル選択： 複数の統計モデルの中から最適なモデルを選ぶ

– より良いモデルの探究は相対評価
– クロスバリデーションは複雑な手法でも適用可能な利便性あり
  ただし,とにかく評価指標の良いモデルを選択

赤池情報量規準によるモデル選択

– 予測性能が同程度であれば, より単純なモデルを選択

AIC = -2 最大対数尤度 + 2 統計モデルの自由パラメータ数

オッカムの剃刀：「必要が無いなら多くのものを定立してはならない。少数の論理でよい場合は多数の論理を定立してはならない」(wikipedia)

ただし, 想定した統計モデル(パラメトリックモデル)の中に真の分布が含まれているという仮定が必要。例えば, 目的変数が2値データの場合に線形回帰モデルを含めたモデル選択には不向き

AICによるモデル選択の例 #1

Rのコード

```
x = seq(0,10,0.1)
set.seed(1111)
y = sin(x) + 0.2*x + 0.8*runif(length(x))
data = data.frame(rbind(y,x))

reg = glm(y~poly(x, degree = 50, raw =
TRUE),data,family="gaussian")
reg$aic

plot(y)
lines(fitted(reg))
```

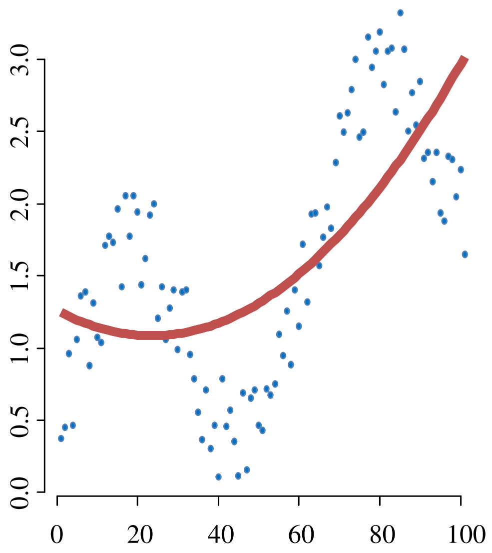
M 次の多項式回帰モデル

– 目的変数の多項式で表現される回帰モデル

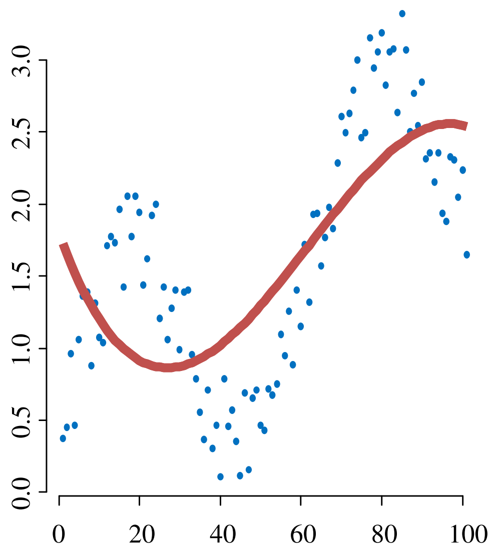
$$y = f(x, \boldsymbol{\theta}) = b_0 + b_1 x + b_2 x^2 + b_3 x^3 +, \cdots, + b_M x^M$$

回帰係数は最小2乗法で推定可能

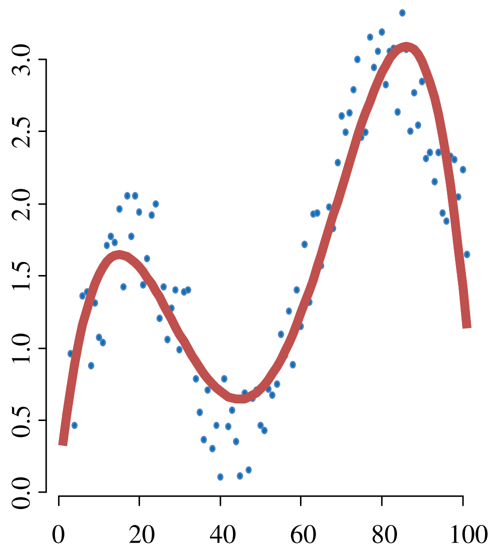
非線形構造をもつデータの傾向に合わせた関数で回帰できる



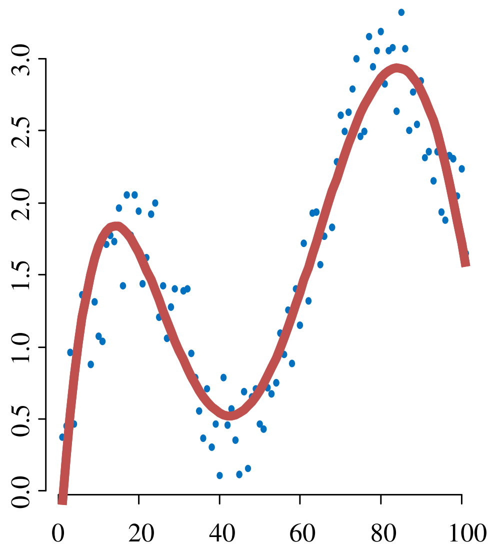
2次の多項式モデル



3次の多項式モデル



4次の多項式モデル

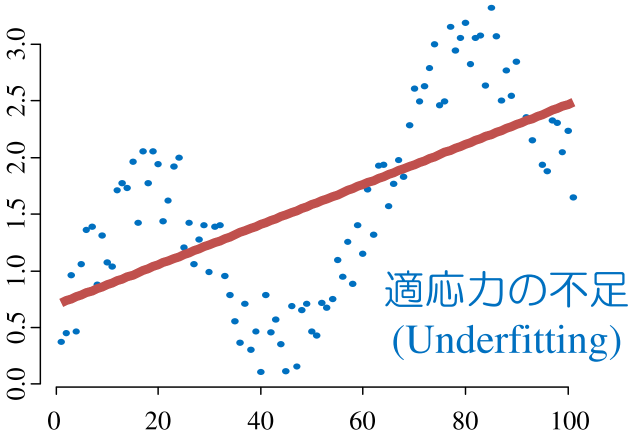


5次の多項式モデル

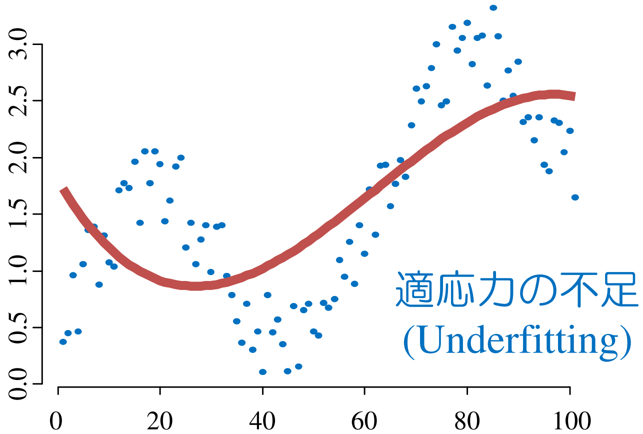
AICによるモデル選択の例 #2

AICを用いることで最適な多項式の次数を選択可能

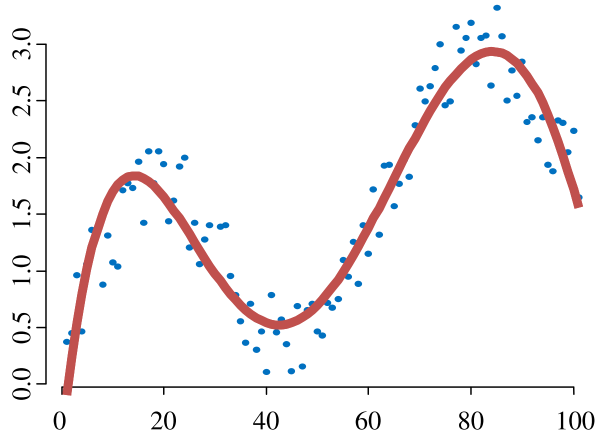
– 複雑な関数による過剰適合（過学習）を回避できる



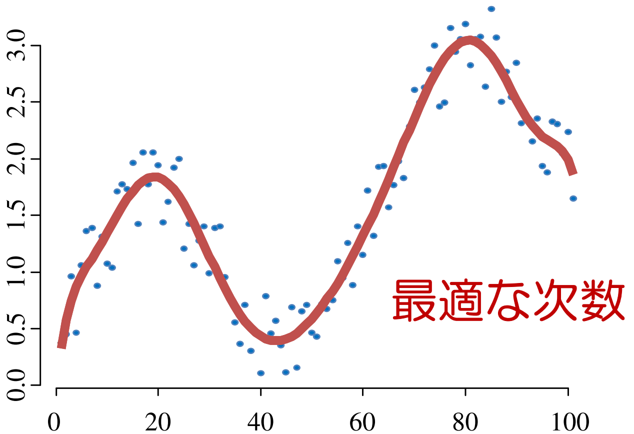
1次の多項式 (AIC = 215.5)



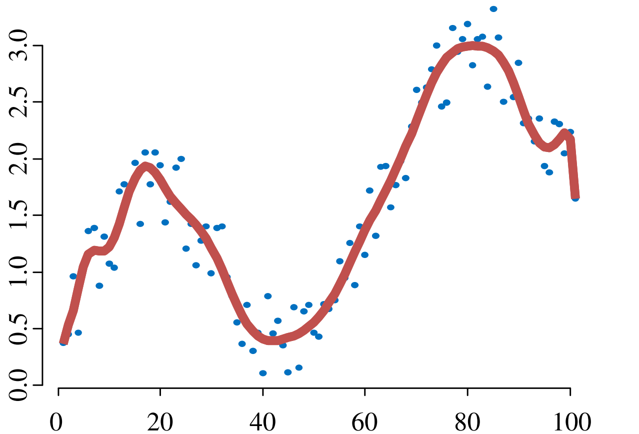
3次の多項式 (AIC = 196.0)



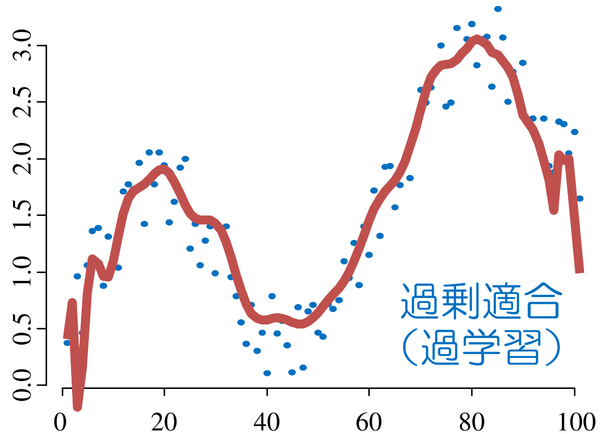
5次の多項式 (AIC = 47.1)



10次の多項式 (AIC = 7.8)



20次の多項式 (AIC = 19.0)



50次の多項式 (AIC = 99.9)

モデルは世界観

伊理正夫, 『モデリング』, *モデリング*（室田, 池上, 土屋編）, 近代科学社, 2015

“モデル(模型)”というからには, それは「本物ではない」ということである。現実そのものではなく, そこに内在する本質的なものを取り出したものである。“取り出す(abstract)”とは“抽象”である。取り出すからにはそこに取り残されたものがあるわけであるから, “抽象”とはすなわち“捨象”である。

抽象の過程は決して純客観的ではありえない。<u>「このことの本質はこれこれである」ということではなくて「私は, これこれがこのことの本質であると見るのだ」という, 主体的な行動である</u>。モデルを作ることは自分の立場の表明であり, 大げさに言えば, 自分の世界観の宣言である。

捨象:（概念を抽象する作用の反面として）現象の特性・共通性以外を問題とせず、考えのうちから捨て去ること。