

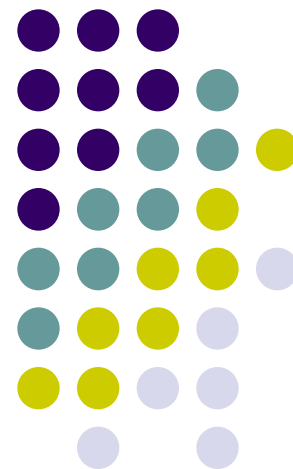
赤池情報量規準

Akaike's Information Criterion

AIC

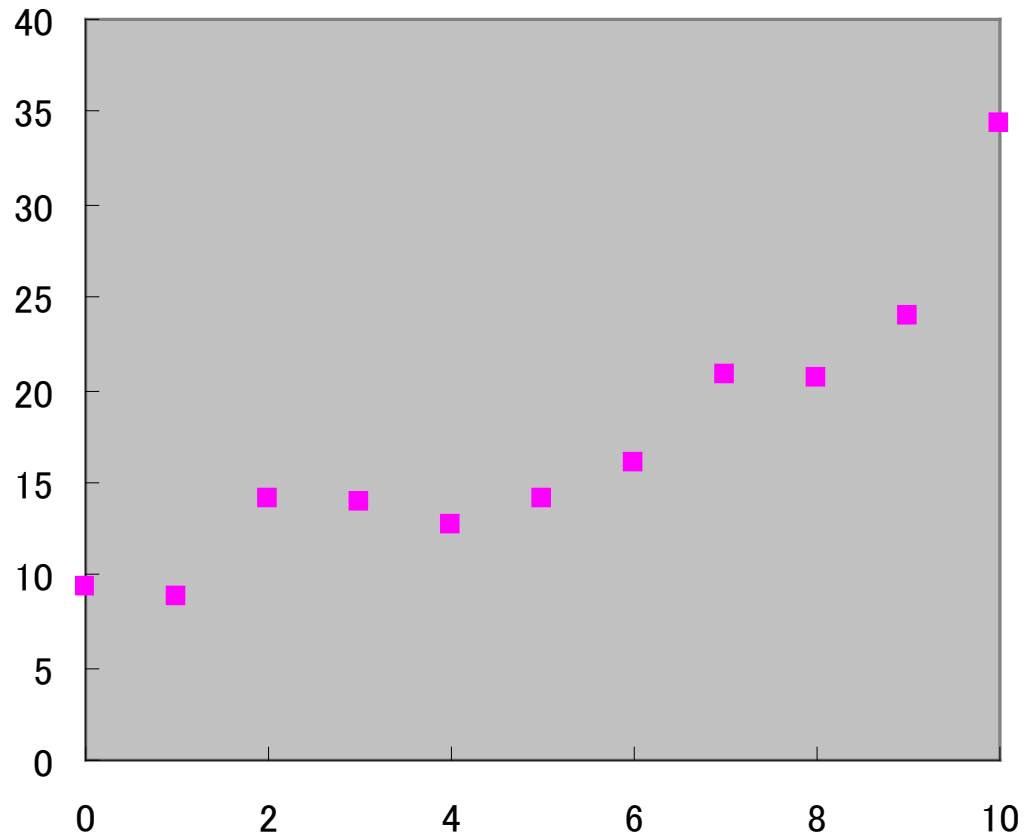
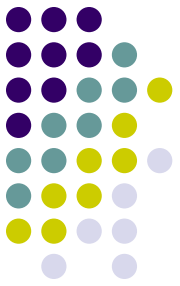
北海道大学大学院水産科学研究院

松石 隆



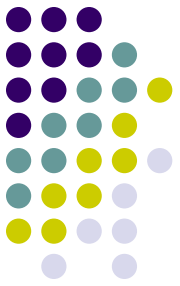
理論曲線のおてはめ

Curve Fitting



手順

curve fitting procedure



1. モデルを決める / decide a model

$$y = a + bx + cx^2$$

2. 初期値を決める / set initial parameters

$$y = 6 + 0.2x + 0.3x^2$$

手順

curve fitting procedure

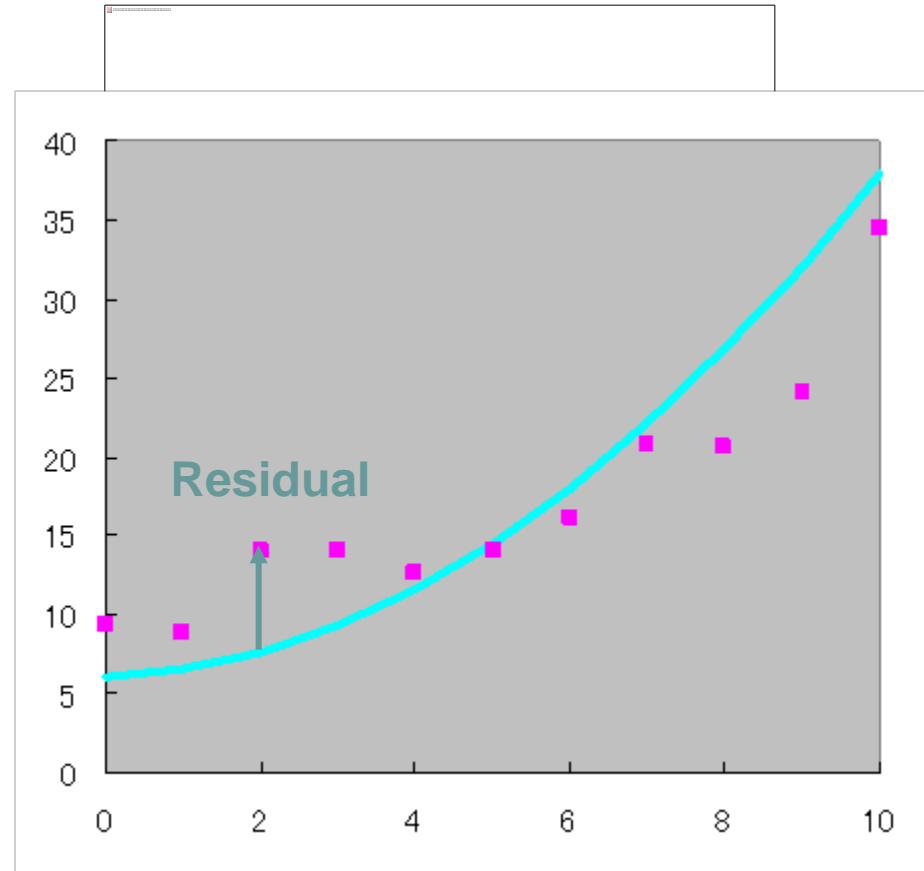


3 残差平方和の計算

sum of square residuals

$$SS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\hat{y}_i = 6 + 0.2x_i + 0.3x_i^2$$



手順

Curve fitting procedure



4 残差平方和が最少になるパラメータを探す

Find the parameters that minimize the SS

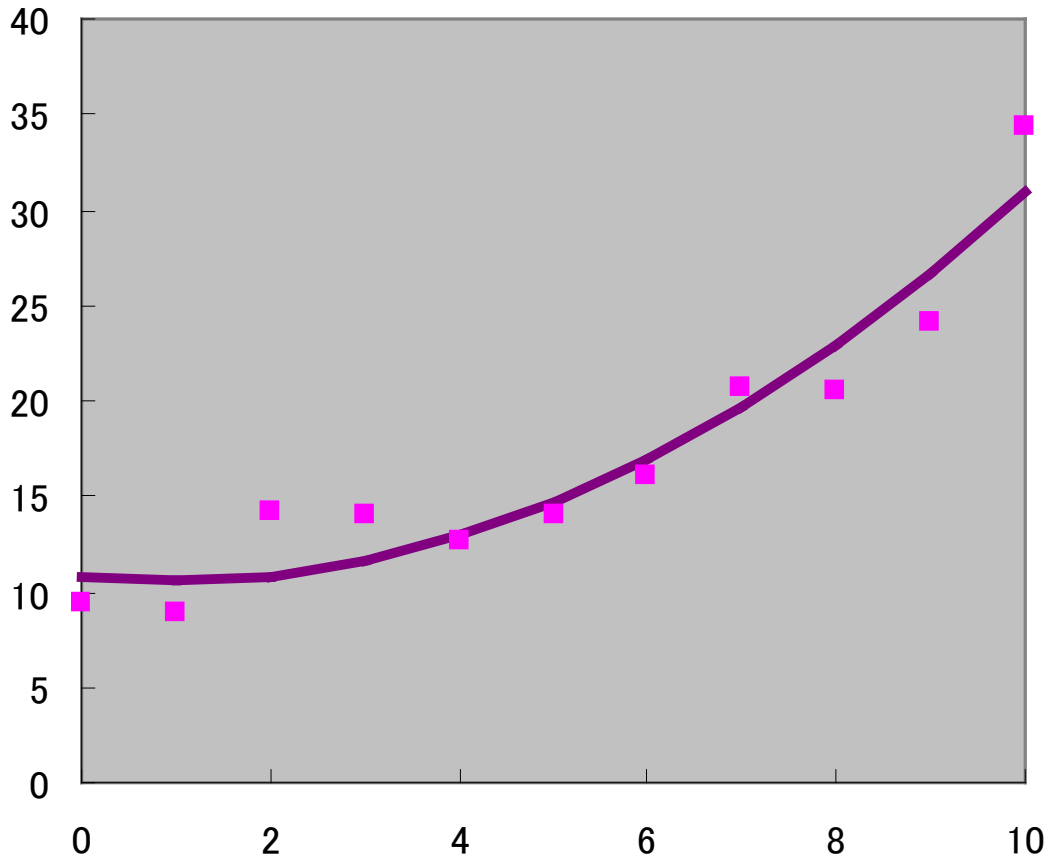
- 微分 differential equations

$$\frac{\partial SS}{\partial a} = 0, \quad \frac{\partial SS}{\partial b} = 0, \quad \frac{\partial SS}{\partial c} = 0$$

- 非線形最適化法 NL optimization method
SOLVER addin (MS EXCEL) etc

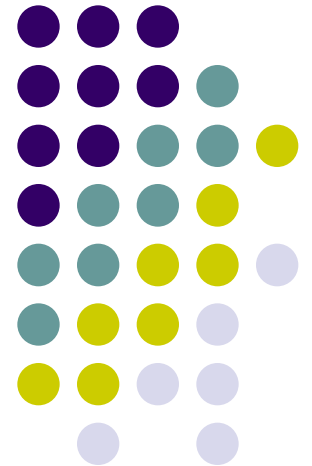


$$\hat{y} = 10.7 - 0.43x + 0.25x^2$$

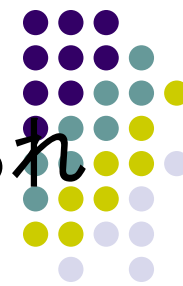


最尤推定量

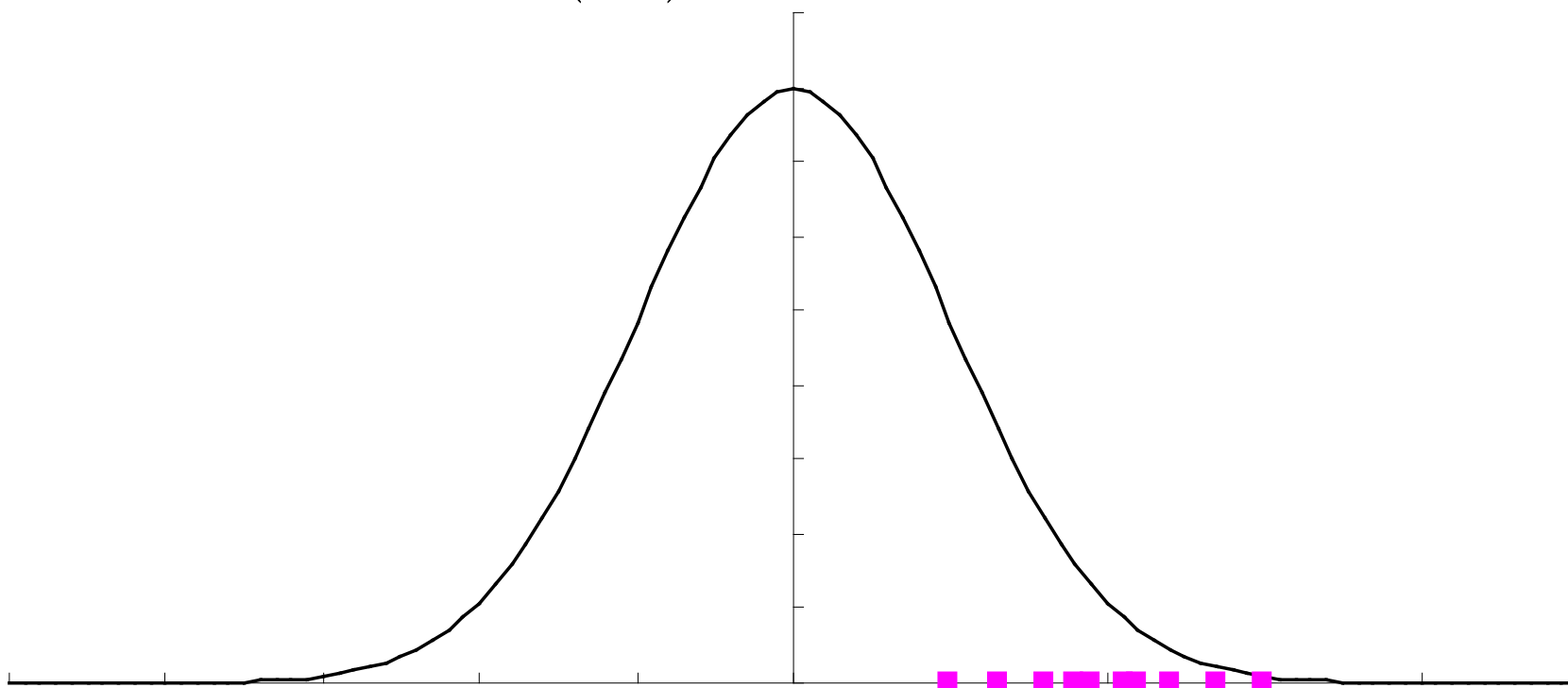
Maximum Likelihood Estimator



- このデータ■は $N(0,1^2)$ に従う母集団から取られた標本のデータだという。尤もらしいか？



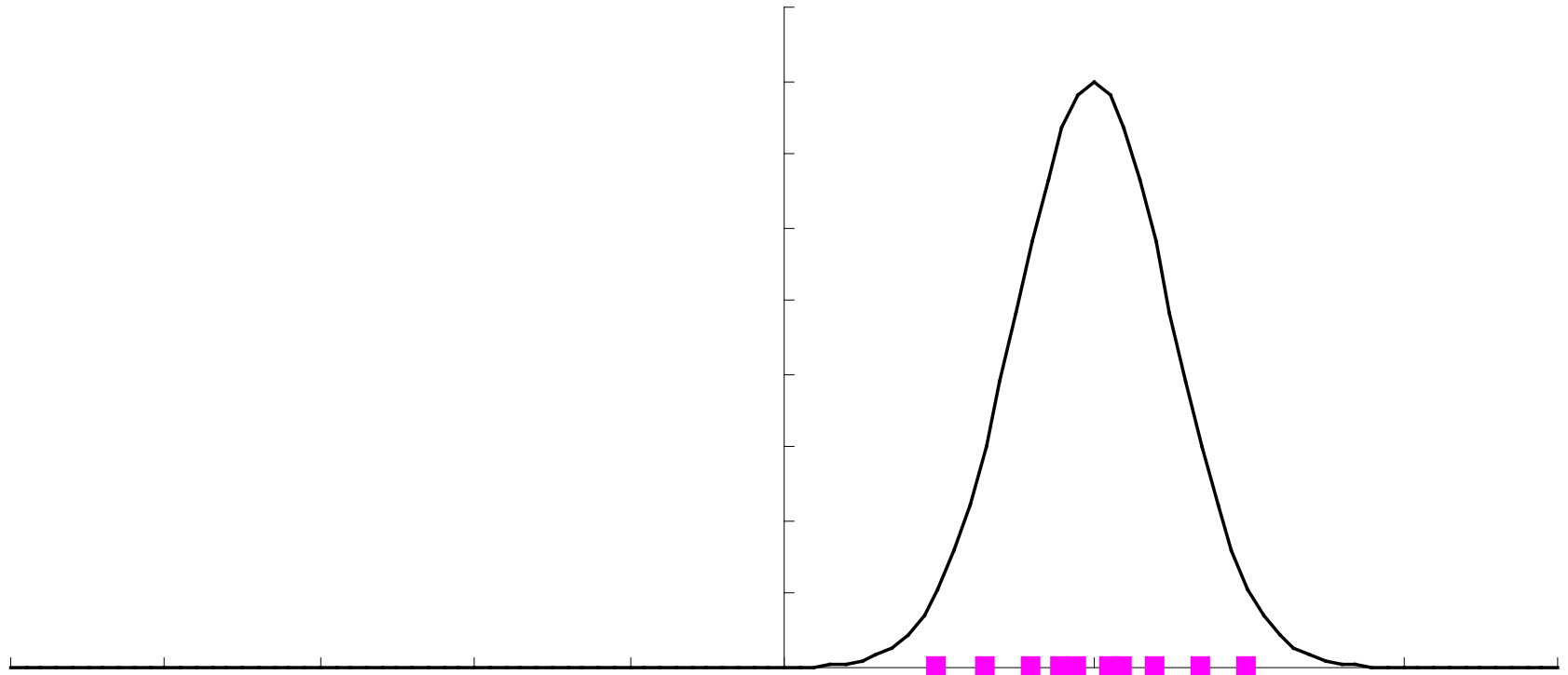
$N(0,1^2)$ の確率密度分布

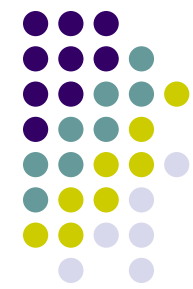


- このデータ■は $N(2, 0.5^2)$ に従う母集団から取られた標本のデータだという。尤もらしいか？

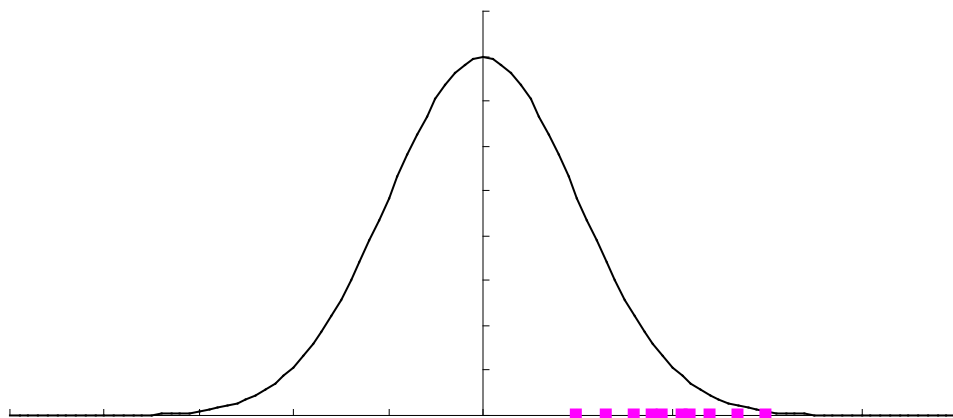


$N(2, 0.5^2)$ の確率密度分布

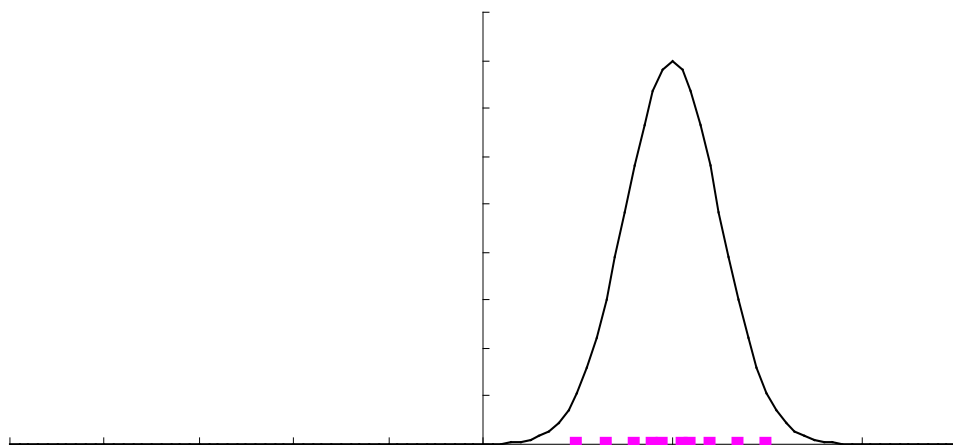




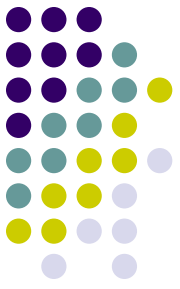
何が尤もらしいのか



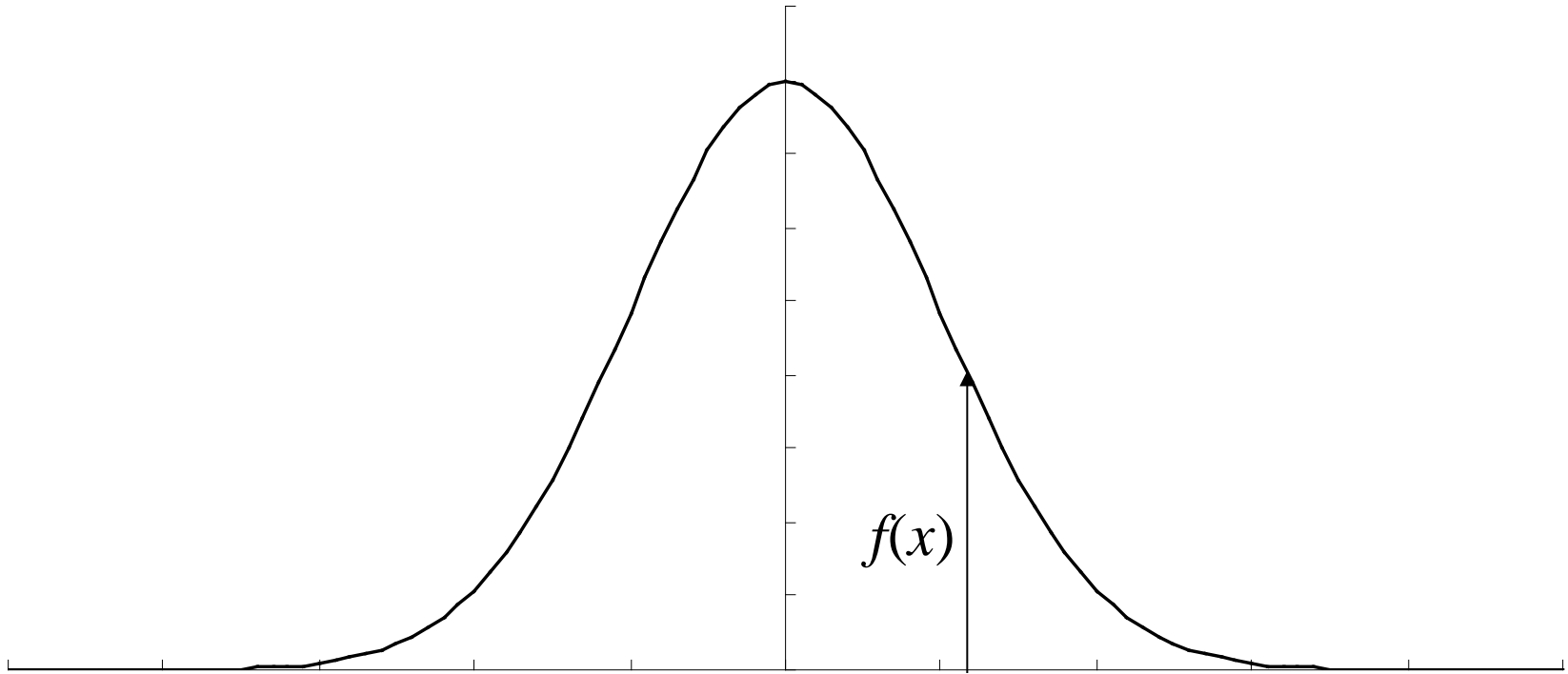
←山が低いところに
データが沢山ある



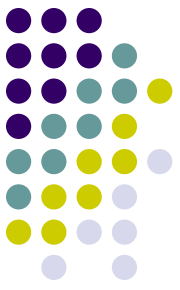
←山が高いところに
データが沢山ある



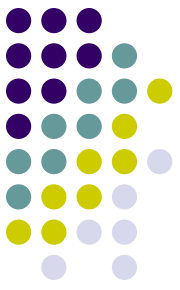
確率密度分布関数の値が、そのデータが出てくる
確率(尤もらしさ)を表している



$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

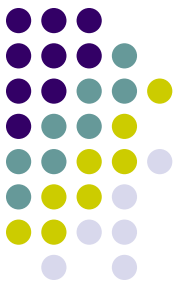


- データ x が生じる確率は $f(x)$ に比例する。
- これを尤度と呼ぶ。
- データ x_1, x_2 が同時に生じる確率は $f(x_1)f(x_2)$ と積に比例するので、その時の尤度を $f(x_1)f(x_2)$ とする。
- 尤もらしさは $f(x)$ を決定するパラメータ(ここでは μ と σ^2)によって変化する



尤度

- もっともらしさ
- 「パラメータ θ をもつ確率密度関数に従う確率分布からデータ x が出た」もっともらしさは $f(x; \theta)$ で表す
- 「パラメータ θ をもつ確率密度関数に従う確率分布からデータ $x_i, i=1, n$ が出た」もっともらしさは $f(x_1, \dots, x_n; \theta)$ で表す

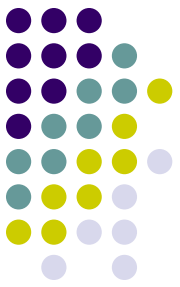


- これを尤度という

$$l(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta)$$

- x_i が互いに独立の場合

$$\begin{aligned} l(\theta; x_1, \dots, x_n) &= f(x_1, \dots, x_n; \theta) \\ &= f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta) \\ &= \prod_{i=1}^n f(x_i; \theta) \end{aligned}$$



最尤推定量

- 最も尤度が高くなる時のパラメータを以て，推定値とする方法
- $\hat{\theta}(X_1, \dots, X_n)$ と表記する
- 尤度が最大になるとき対数尤度 $\log l(\theta; x_1, \dots, x_n)$ も最大になることから，対数尤度を用いて，最尤推定量を計算するともある。

例：正規分布

Example Normal Distribution



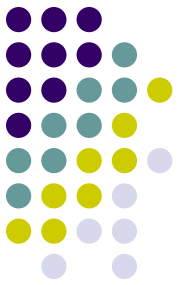
$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$$

$$\ln f(x | \mu, \sigma^2) = -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (x - \mu)^2$$

$$LL = \sum_{i=1}^n \ln f(x_i | \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

データ数とパラメータが与えられると決まる値

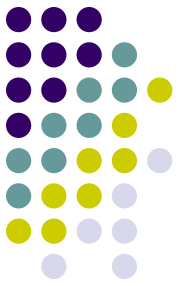
μ からの
残差平方和



ではどうやってモデルを決めるのか？
How do you decide the model?

候補となるモデル

candidate polinomials



$$\hat{y} = c_0$$

$$\hat{y} = c_0 + c_1x$$

$$\hat{y} = c_0 + c_1x + c_2x^2$$

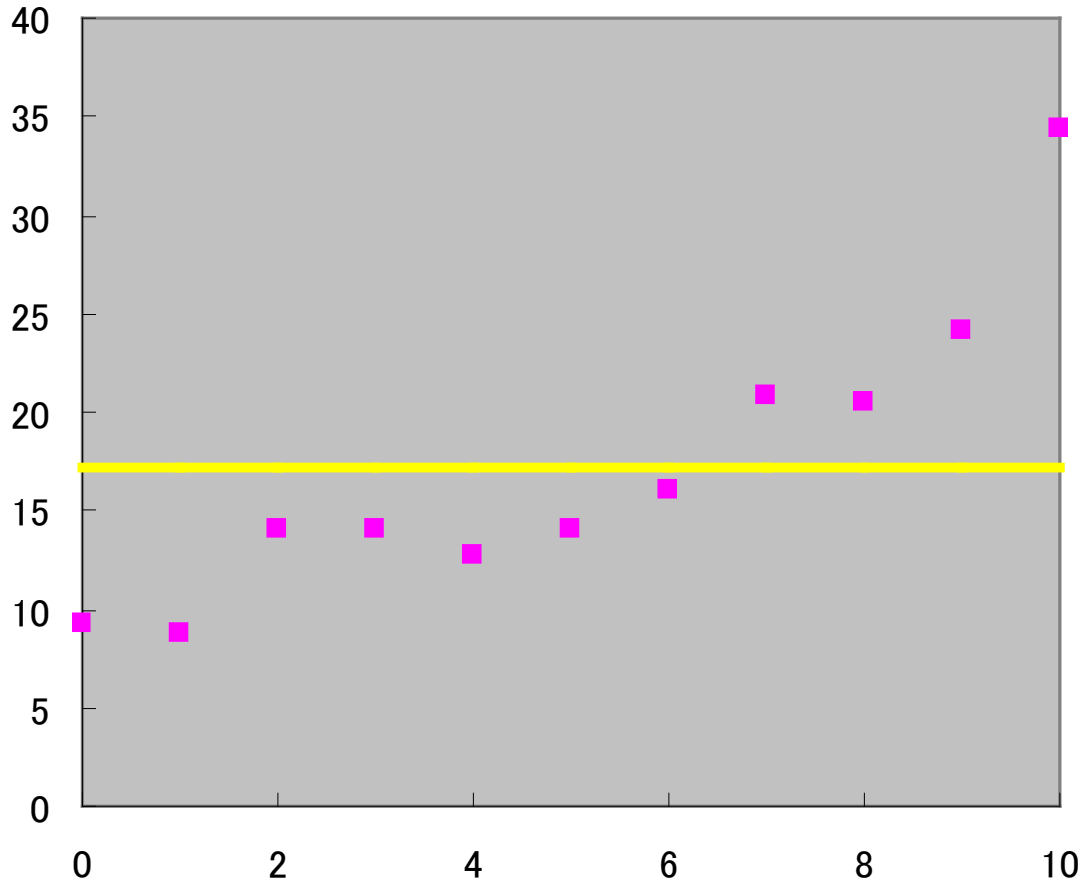
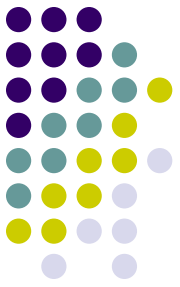
$$\hat{y} = c_0 + c_1x + c_2x^2 + c_3x^3$$

$$\hat{y} = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4$$

⋮

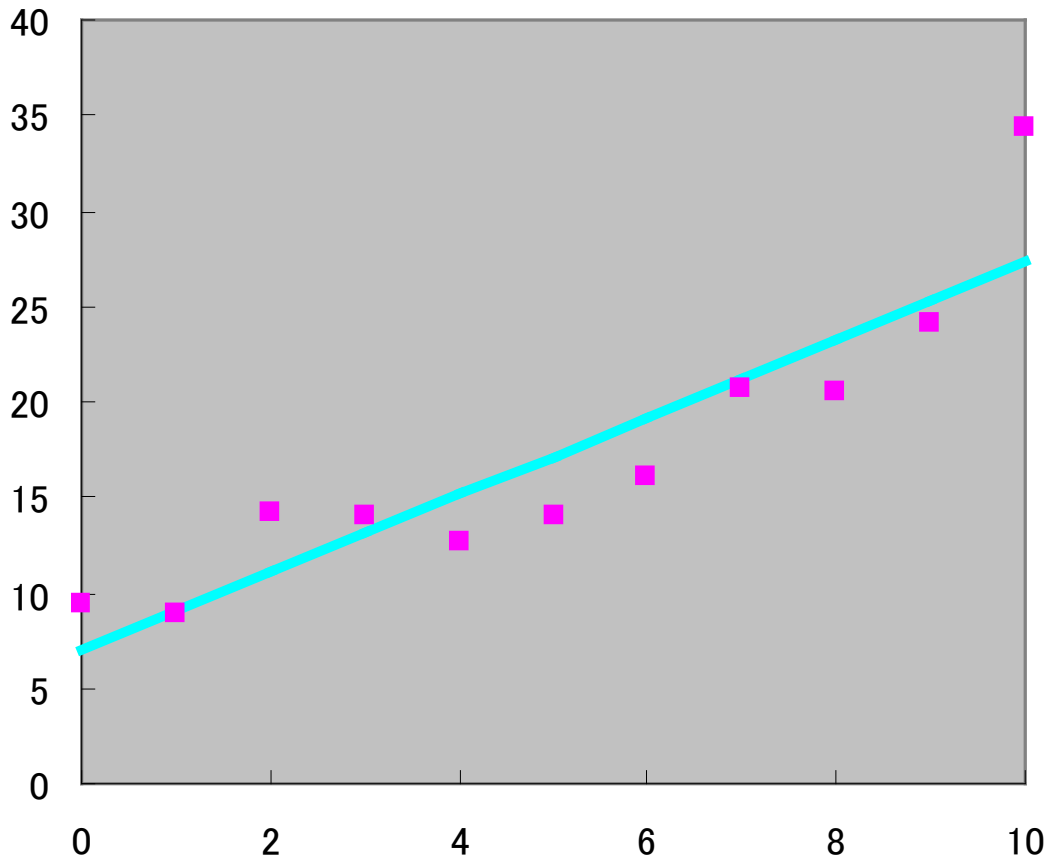
$$\hat{y} = \sum_{i=0}^m c_i x^i$$

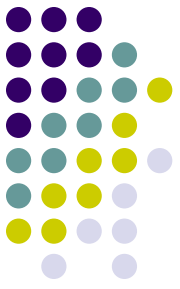
$$\hat{y} = c_0$$



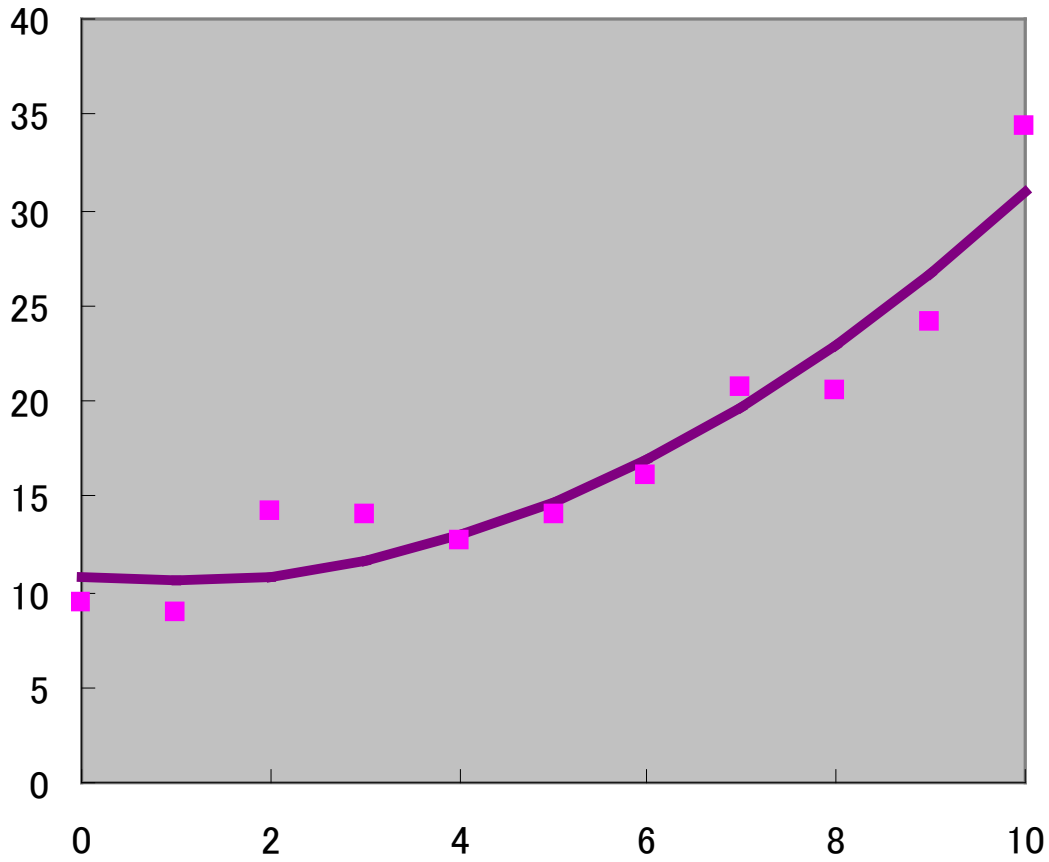


$$\hat{y} = c_0 + c_1x$$



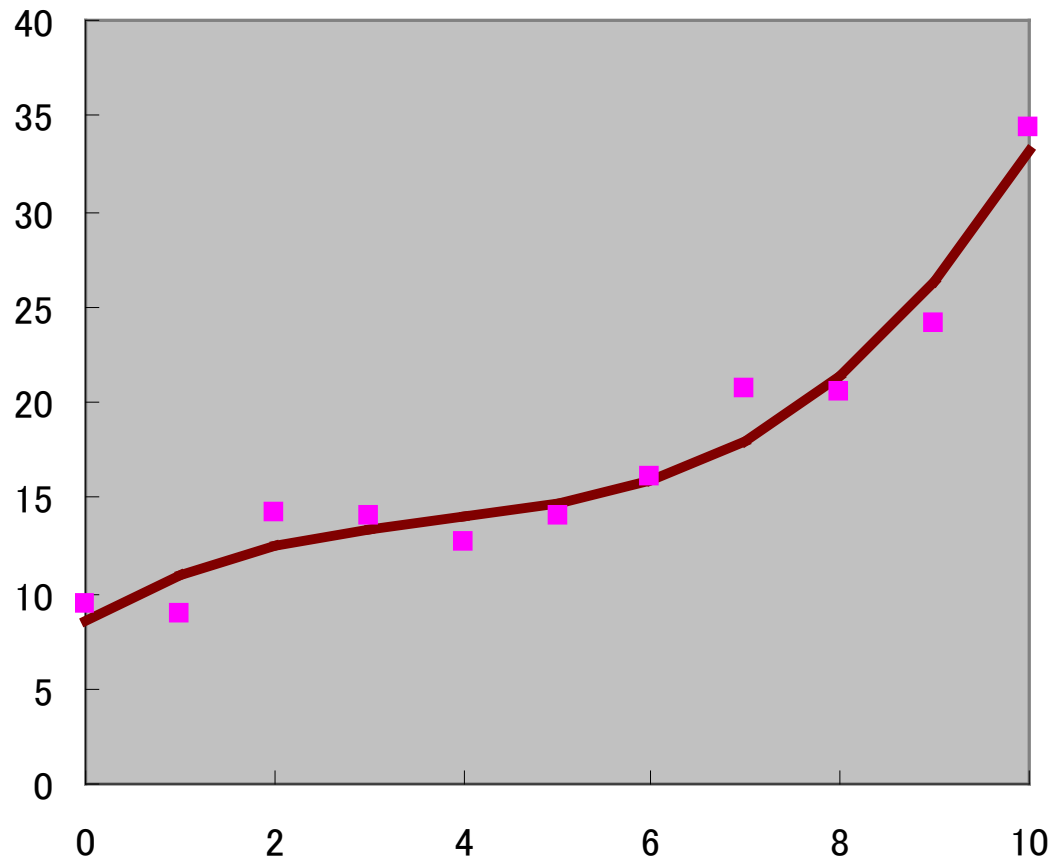


$$\hat{y} = c_0 + c_1x + c_2x^2$$



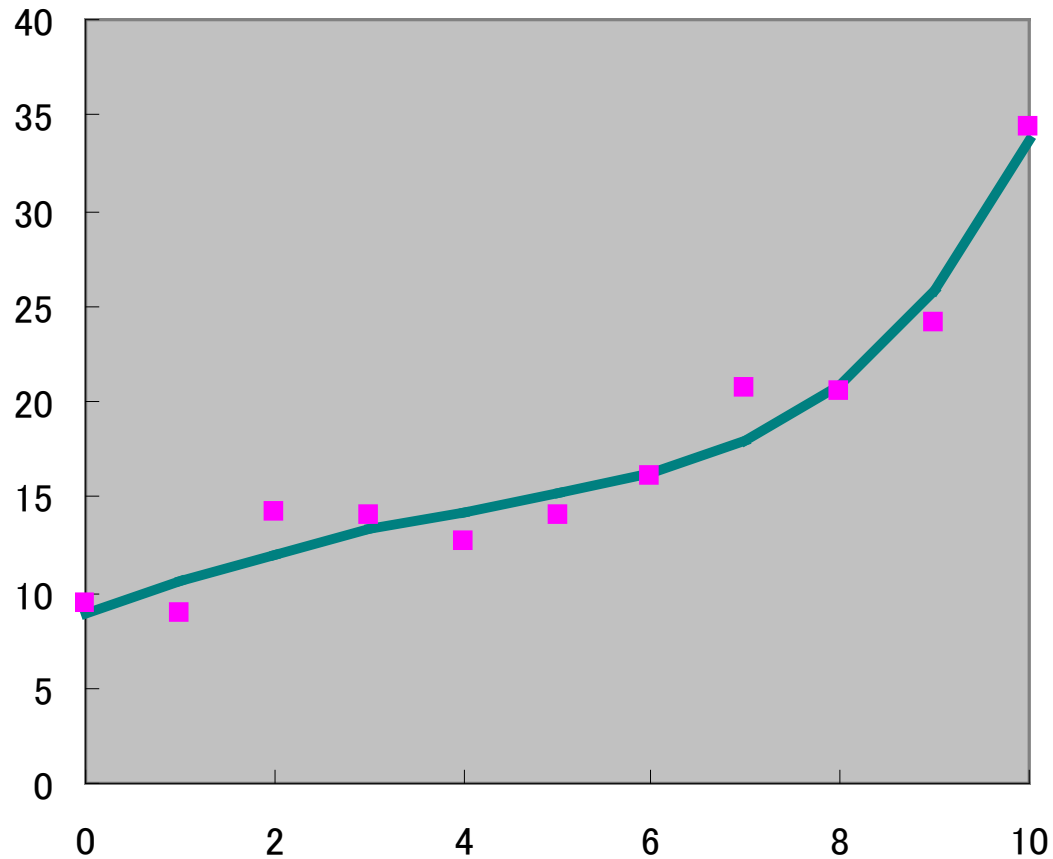


$$\hat{y} = c_0 + c_1x + c_2x^2 + c_3x^3$$

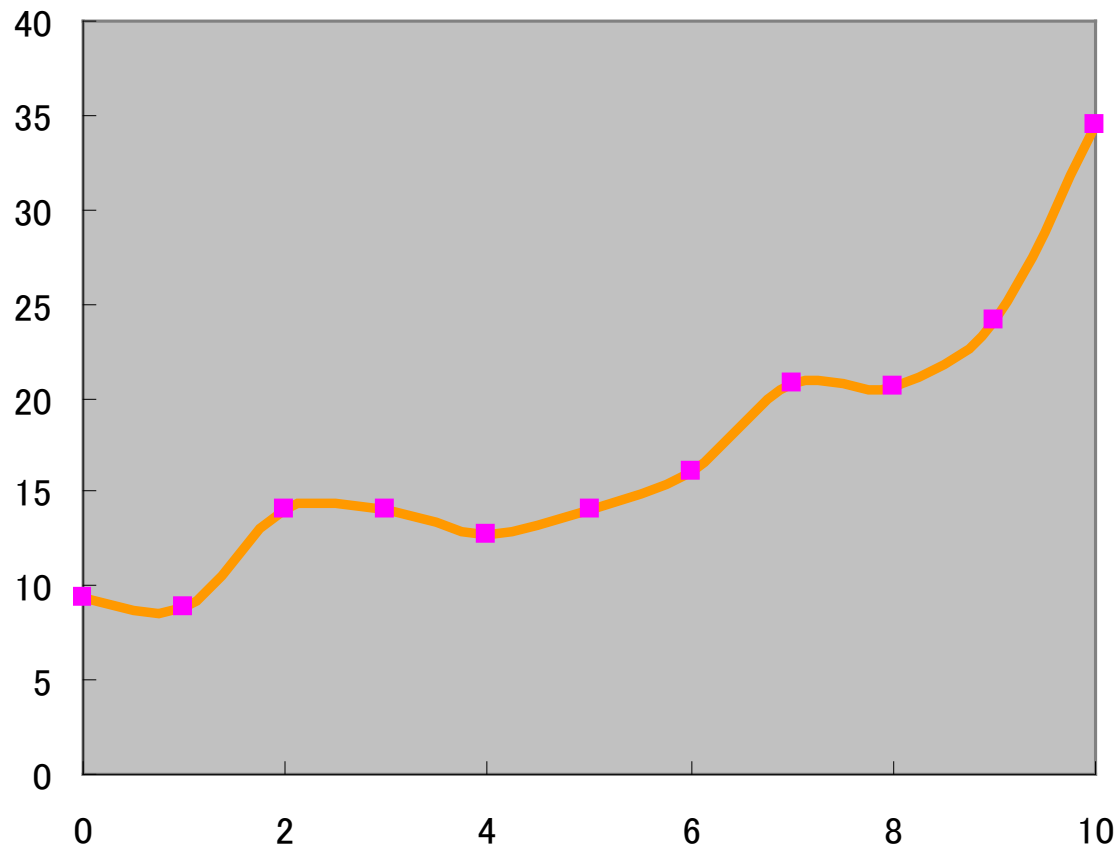
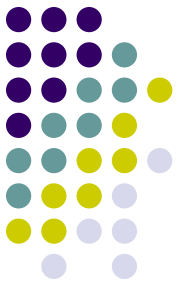




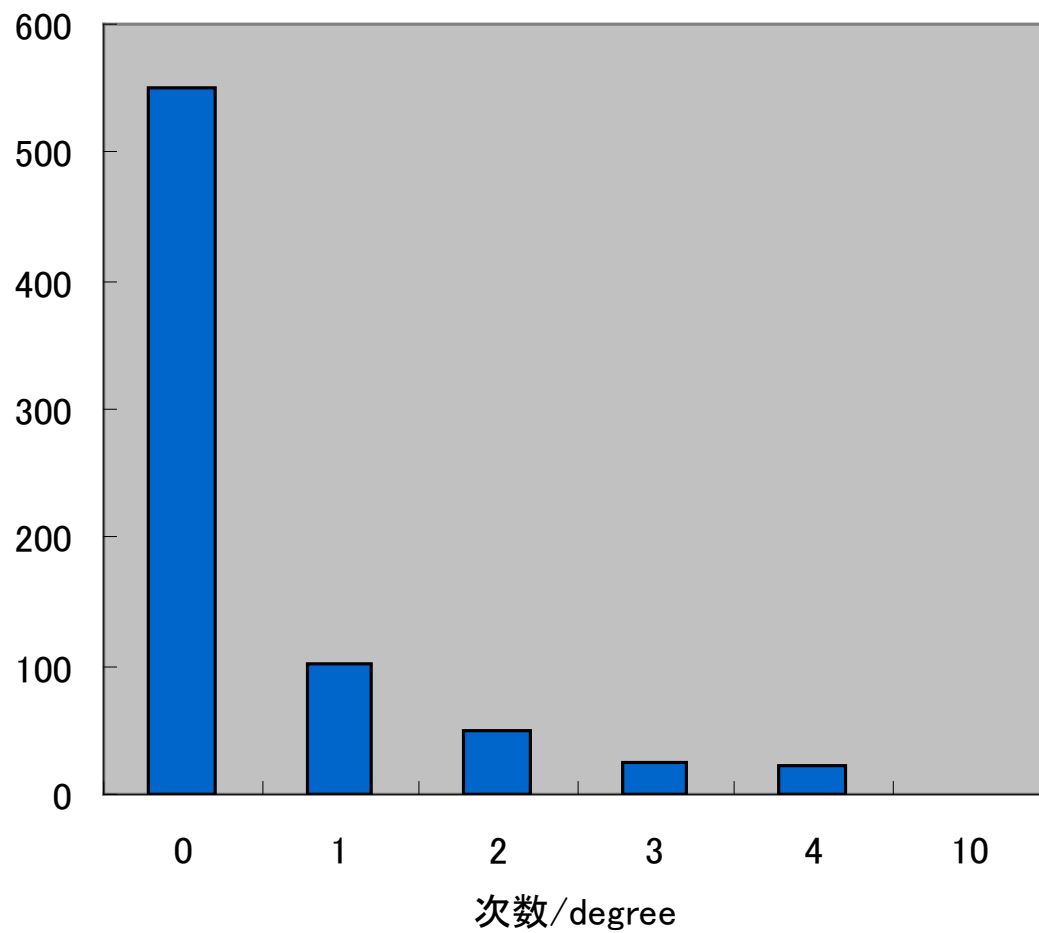
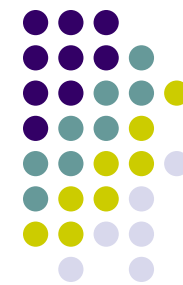
$$\hat{y} = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4$$



$$\hat{y} = \sum_{i=0}^{10} c_i x^i$$



残差平方和 SS



当てはまりと次数

Fitting and degree of polynomials



- 次数が高いほど残差平方和は小さい

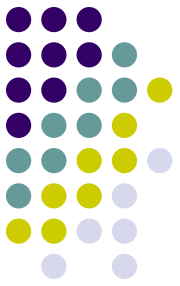
The higher the degree is, the smaller the SS is

$$SS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- 次数がデータ数と等しくなると完璧にデータに当てはめる事が可能

Theoretically, when the degree is equal to the number of data, the curve is perfectly fit the data.

$$\hat{y} = \sum_{i=1}^n c_i x^i = \prod_{i=1}^n (x_i - x)$$

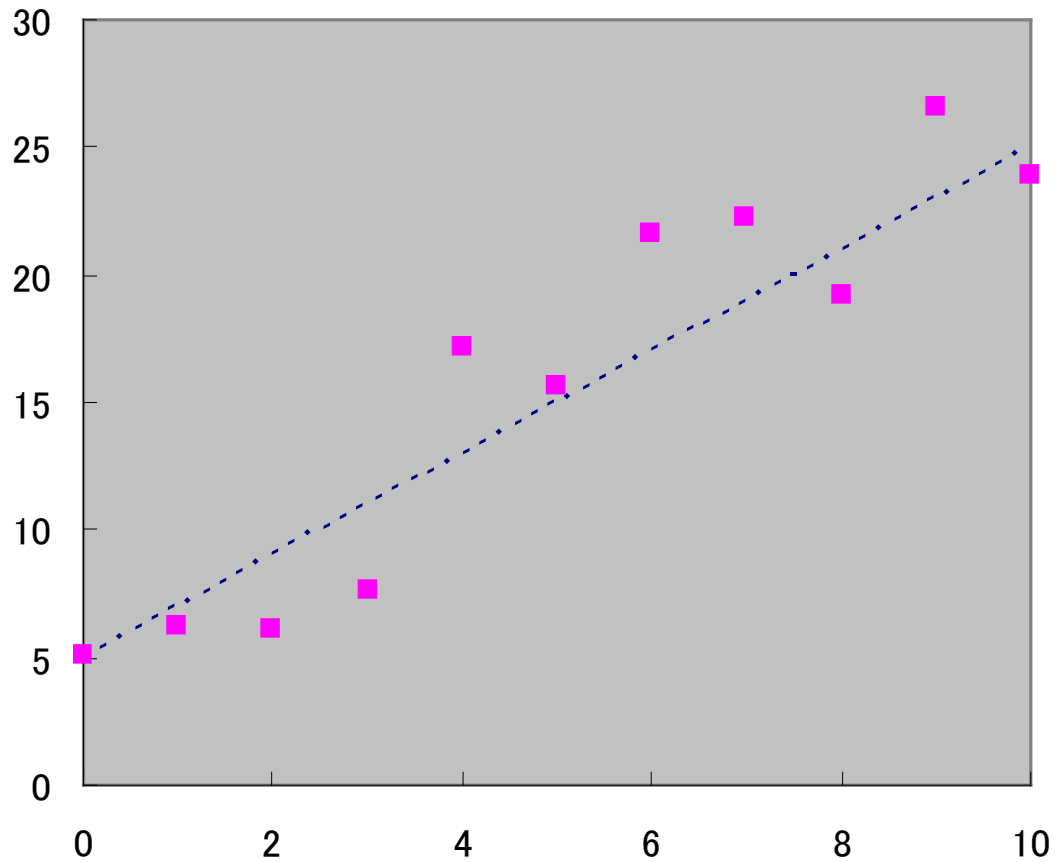


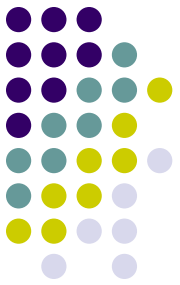
次数が多ければ多いほど良いのか

Really the more the better?

別の乱数を振る

Make other data from the same liner equation

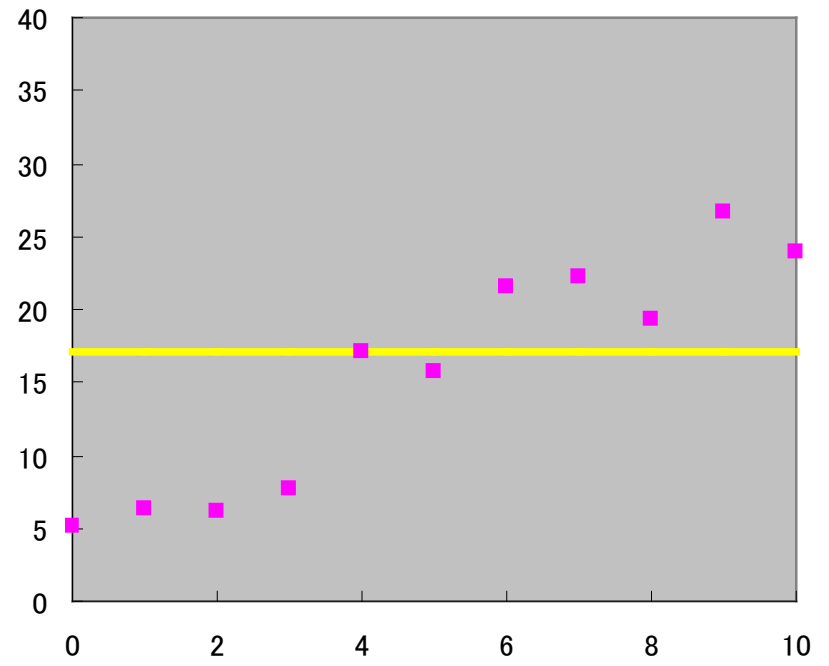
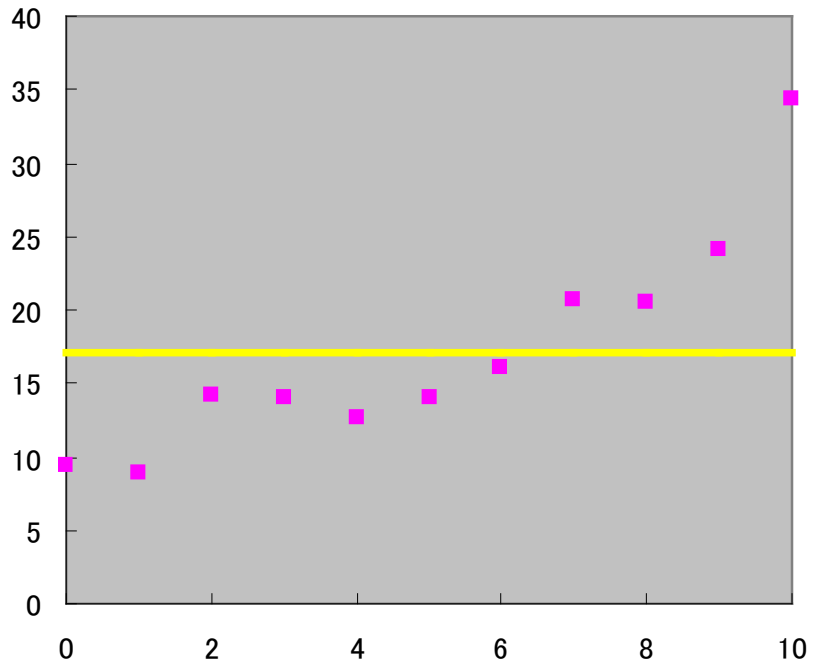




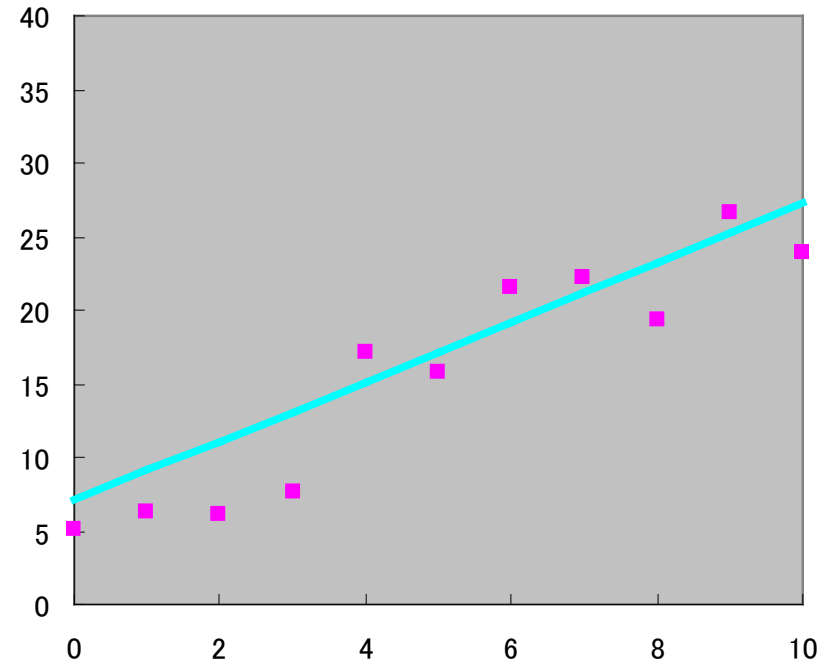
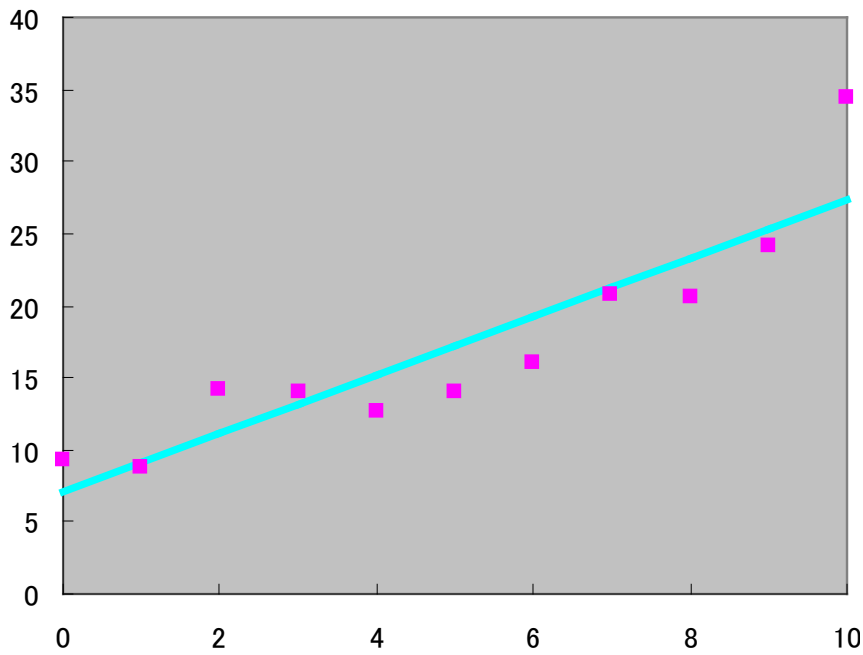
**先ほど推定したモデルは
新しいデータを予測できているか？**

**Does the estimated models predict
the new data?**

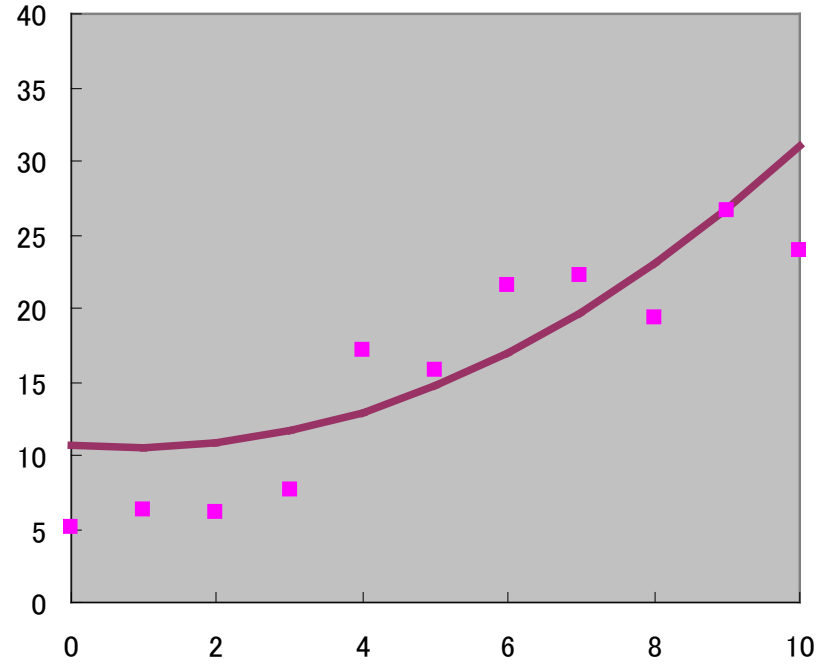
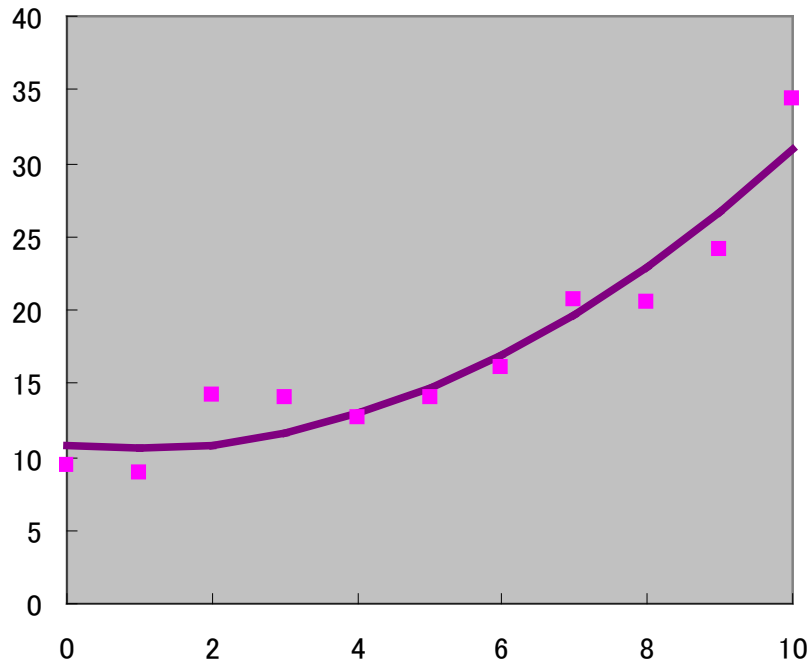
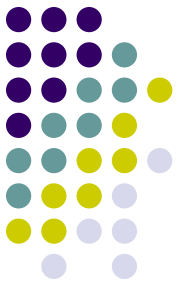
$$\hat{y} = c_0$$



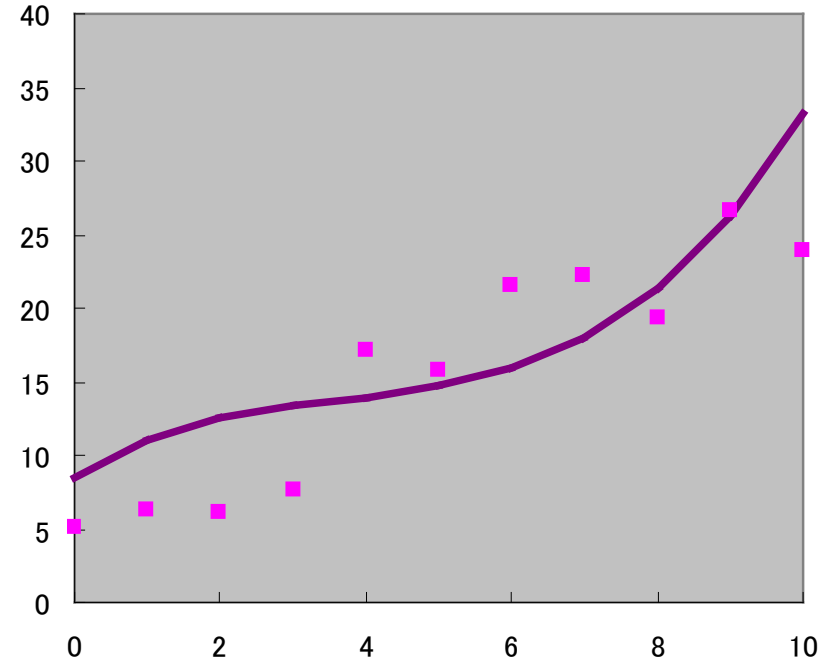
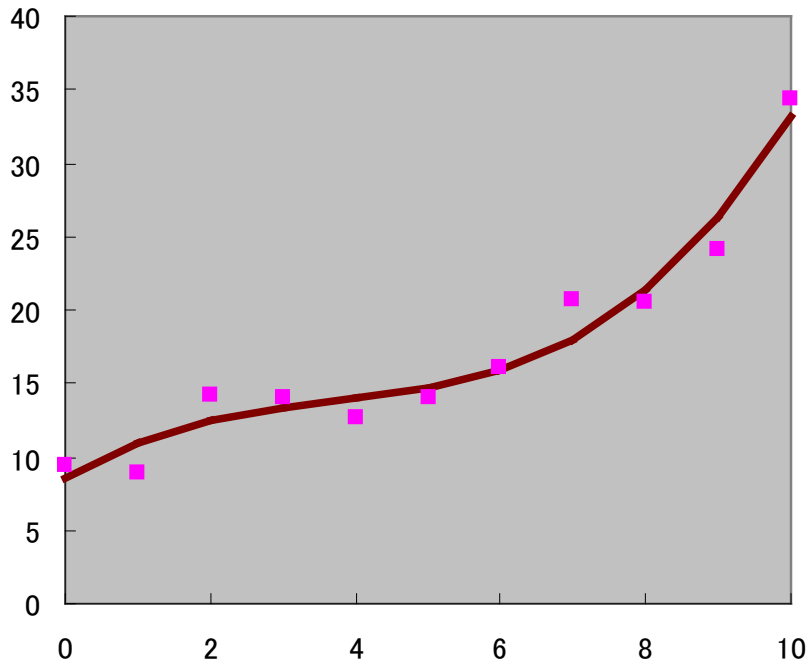
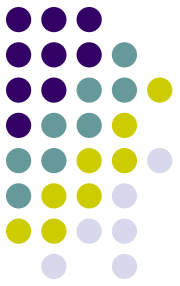
$$\hat{y} = c_0 + c_1x$$



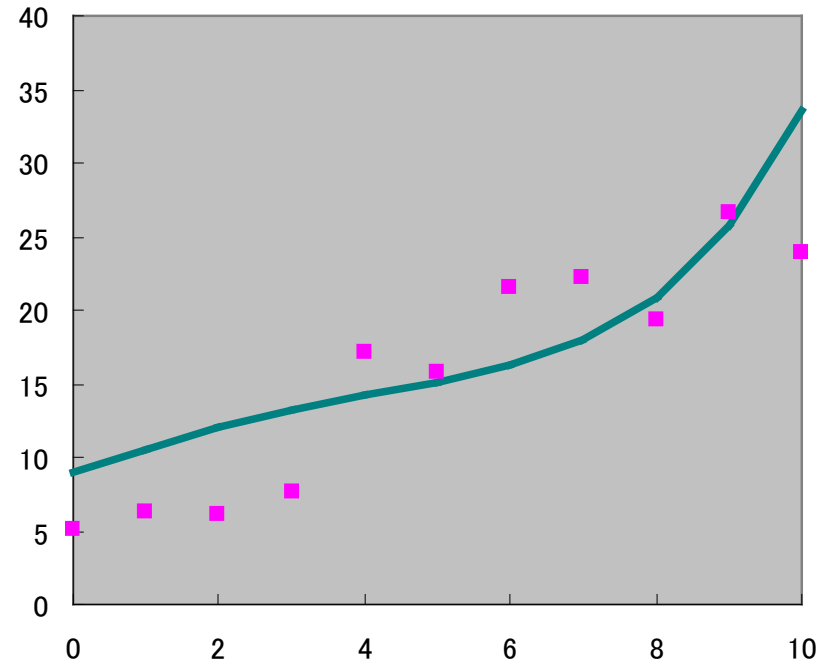
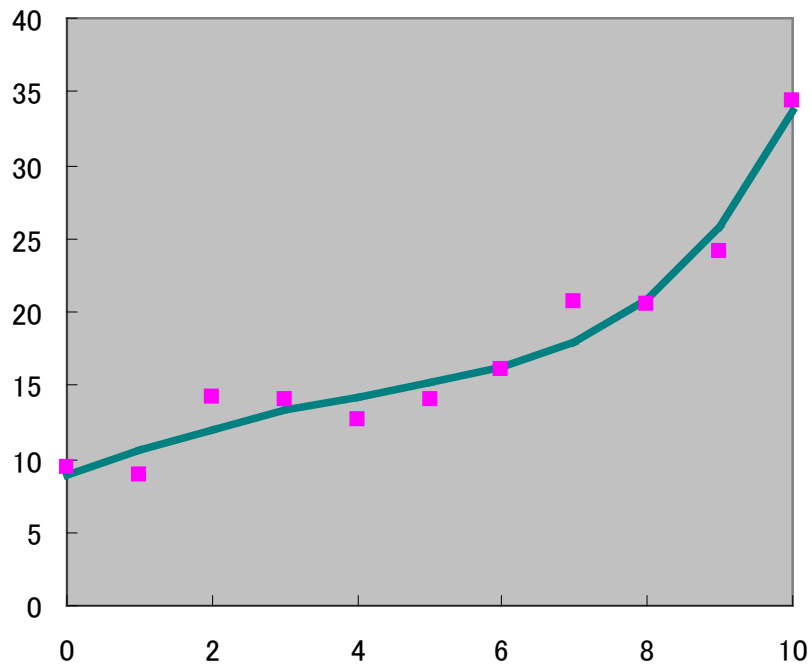
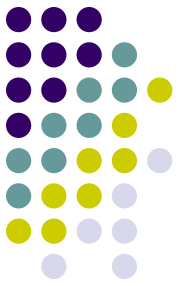
$$\hat{y} = c_0 + c_1x + c_2x^2$$



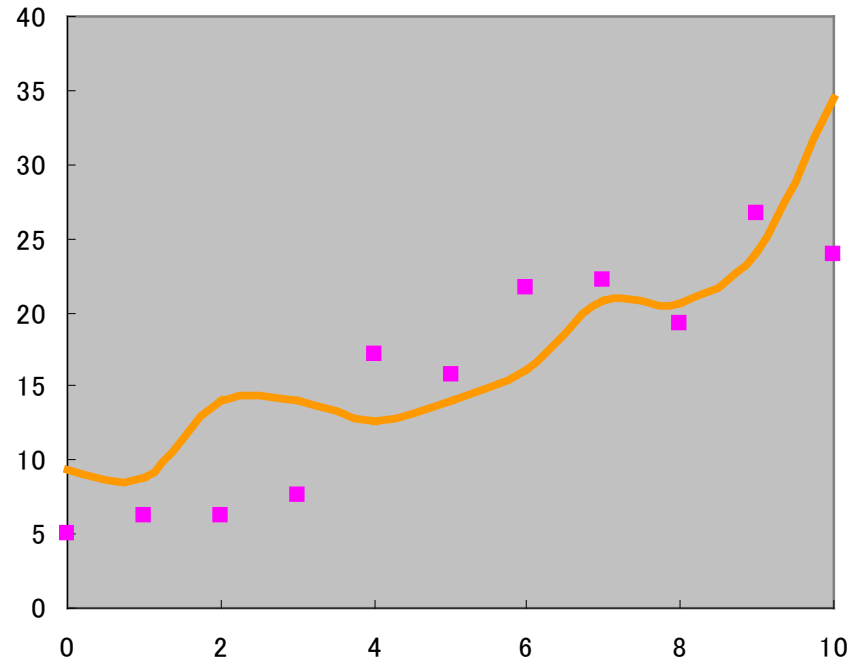
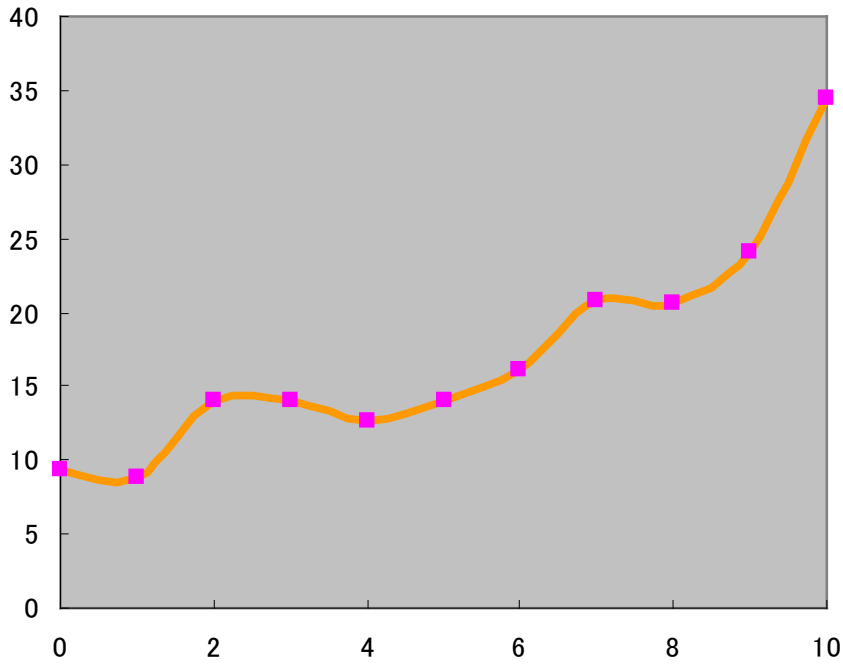
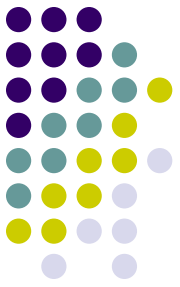
$$\hat{y} = c_0 + c_1x + c_2x^2 + c_3x^3$$



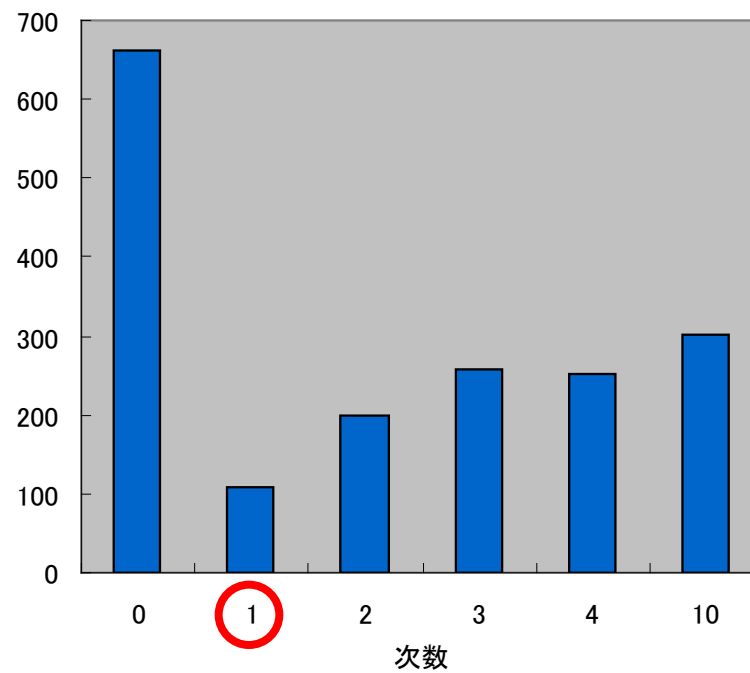
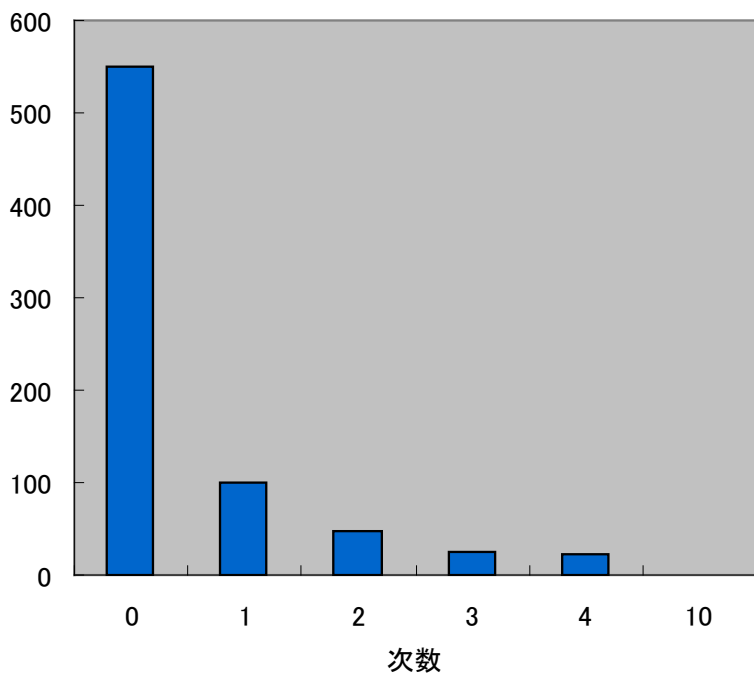
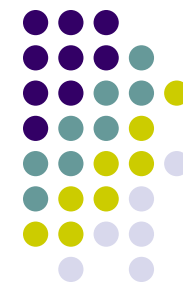
$$\hat{y} = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4$$

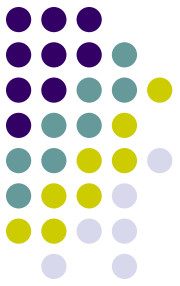


$$\hat{y} = \sum_{i=0}^{10} c_i x^i$$



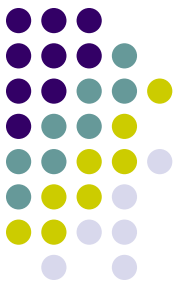
残差平方和 SS





**正しい次数のモデルは
未知のデータにも当てはまりが良い**

**The model with true degree
fit to unknown data.**

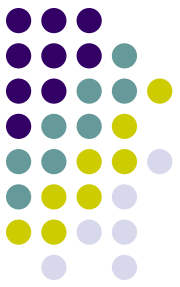


手持ちのデータから正しいモデルを選択する方法

How to select the best model only from the observed data

モデル選択の指標

Criterion for model selection



- 次数が高すぎるモデル high degree
 - 当てはまりが良くなる good fit
 - 未知のデータに当てはまらない bad forecast
- 次数が低すぎるモデル low degree
 - 当てはまりが悪くなる bad fit
 - 未知のデータに当てはまる good forecast
- 当てはまりの指標とモデルの単純さの指標の組み合わせ
Combination of fitting and model simplicity index

AIC 赤池情報量規準

Akaike's Information Criterion



$$AIC(M) = -2 \times MLL(M) + 2 \times k$$

M : モデル model

MLL : 最大対数尤度 maximum log likelihood

k : パラメータ数 number of parameters

正規分布の場合

Example of Normal Distribution

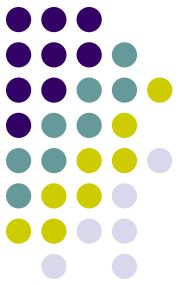


$$AIC(M) = -2 \times MLL(M) + 2 \times k$$

$$MLL = -\frac{n}{2} \ln(2\pi\sigma^{2*}) - \frac{1}{2\sigma^{2*}} \sum_{i=1}^n (x_i - \mu^*)^2$$

$$AIC = n \ln(2\pi) + n \ln \frac{SS}{n} + n + 2(m + 2)$$

利用例



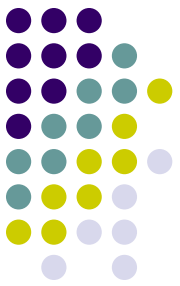
アワビの成長曲線について

Growth Curve of Abalone

松石 隆・斉藤和敬・菅野泰次

北大水産彙報 46: 53-62, 1995

アワビの年齢成長データを成長曲線に当てはめた



1. von Bertalanffy の成長曲線

$$l_{i,t} = L_{\infty}(1 - e^{-K(t-c)}) + \varepsilon_{i,t}$$

2. Logistic 成長曲線

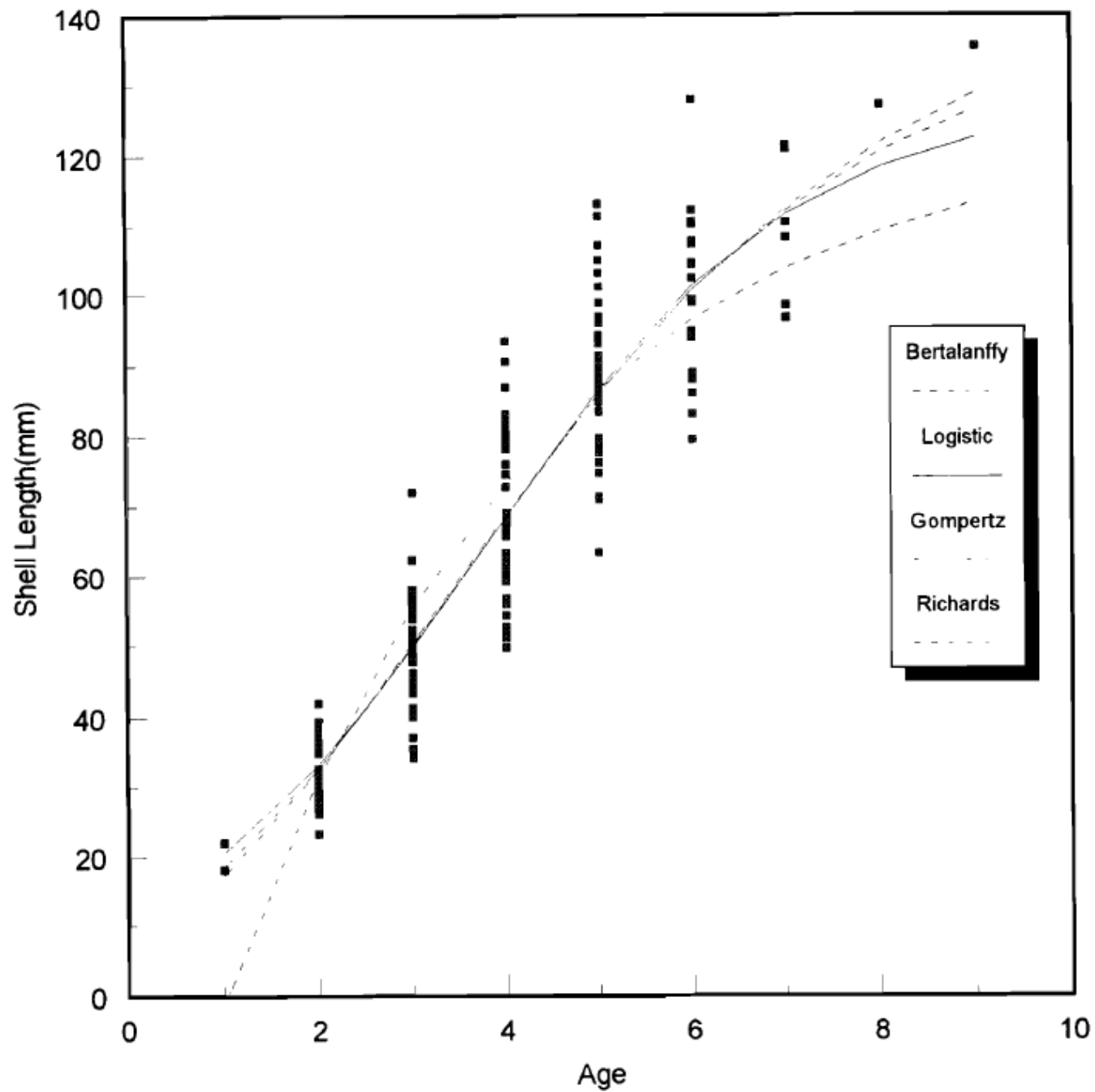
$$l_{i,t} = \frac{L_{\infty}}{1 + e^{-K(t-c)}} + \varepsilon_{i,t}$$

3. Gompertz の成長曲線

$$l_{i,t} = L_{\infty}e^{-ce^{-Kt}} + \varepsilon_{i,t}$$

4. Richards (Richards, 1959) の成長曲線

$$l_{i,t} = [L_{\infty}^{1-m} - (L_{\infty}^{1-m} - c^{1-m})e^{-K(1-m)t}]^{\frac{1}{1-m}} + \varepsilon_{i,t}$$



結果

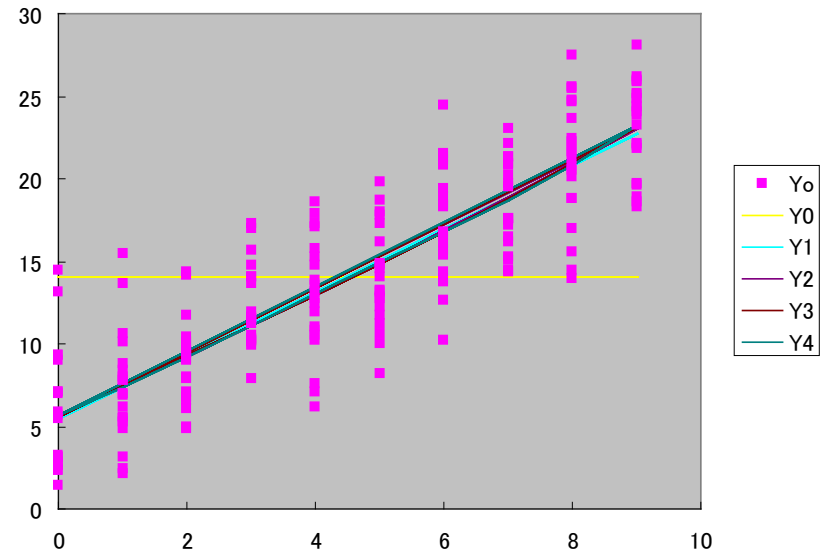
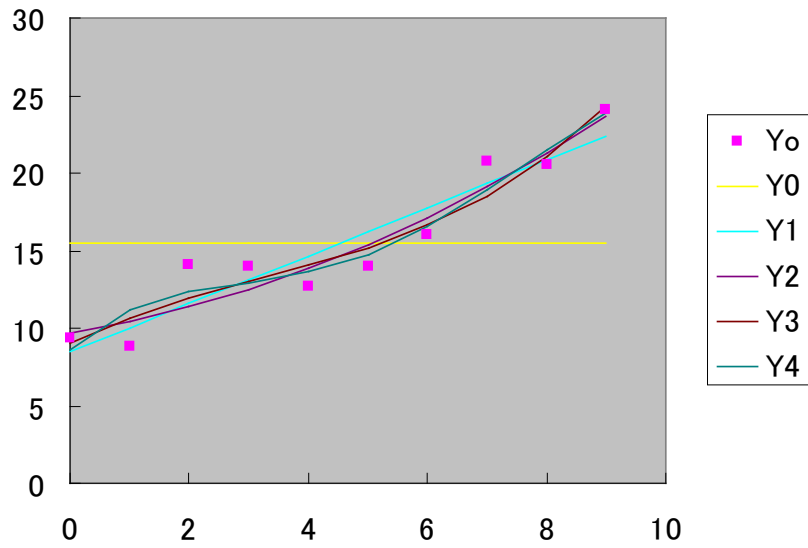
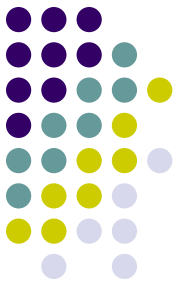


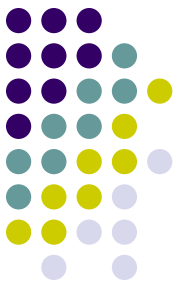
Growth function	L_{∞}	K	c	m	σ_c^2	LL	n	AIC
von Bertalanffy	125.0	0.2976	1.031		79.27	724.68	4	1457.36
(s.e.)	(4.52)	(0.0221)	(0.770)		(7.81)			
Logistic	127.8	0.5982	3.760		99.48	609.90	4	1227.80
(s.e.)	(6.96)	(0.0534)	(0.213)		(10.98)			
Gompertz	147.6	0.3444	3.022		99.21	609.68	4	1227.36★
(s.e.)	(12.08)	(0.0428)	(0.205)		(10.95)			
Richards	139.0	-1.351	9.165	1.313	99.10	609.59	5	1229.18
(s.e.)	(22.8)	(3.232)	(5.462)	(0.702)	(0.11)			

数值例



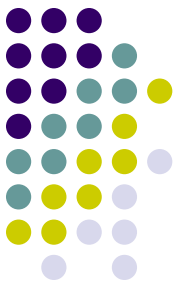
n=10,200





様々なデータ数におけるAIC

Degree	0	1	2	3	4
n=10	63.5	43.9	<u>43.5</u>	44.3	45.4
20	131.9	106.2	<u>100.4</u>	101.0	103.0
50	331.1	264.3	<u>262.7</u>	264.3	266.0
100	661.9	<u>513.6</u>	514.0	515.9	517.7
200	1,314.0	<u>1,040.8</u>	1,042.3	1,044.2	1,046.1



様々な情報量規準

- AIC

$$AIC = -2LL + 2p$$

- c-AIC (有限項補正)

$$c-AIC = -2LL + 3p/2$$

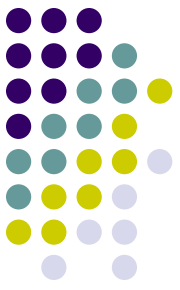
- BIC (ベイズ統計)

$$BIC = -2LL + p \ln(n)$$



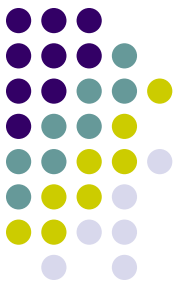
選択された次数

	TRUE	AIC	c-AIC	BIC
n=10	1	2	2	2
20	1	2	2	2
50	1	2	2	1
100	1	1	2	1
200	1	1	1	1



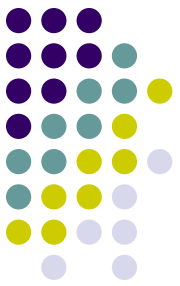
まとめ

- パラメータ数の多いモデルほど、当てはまりが良い
- 当てはまりが良いモデルが良い予測をする訳ではない。
- よい予測をするモデルを選択するためにAICが有用
- AICは相対的な大小のみが意味がある
- データ数がある程度大きい必要がある



Conclusion

- Higher degree model fit to data well.
- Good fit model does not always forecast well.
- AIC is a good tool to choose a better forecasting model.
- Only the relative scale of AIC has mean.
- AIC does not work for small sample.



参考図書

- 赤池弘次／室田一雄著『赤池情報量規準AICモデリング・予測・知識発見』 共立出版 (2007)[[詳細](#)]
- 鈴木義一郎『情報量規準による統計解析入門』 講談社(1995)[[詳細](#)]
- 坂元慶行ほか『情報量統計学』 共立出版 (1983)[[詳細](#)]