# 30分だけでは決してよくわからない
# とてもとても難しい
# 一般化線形モデル with R

M1 白砂優希

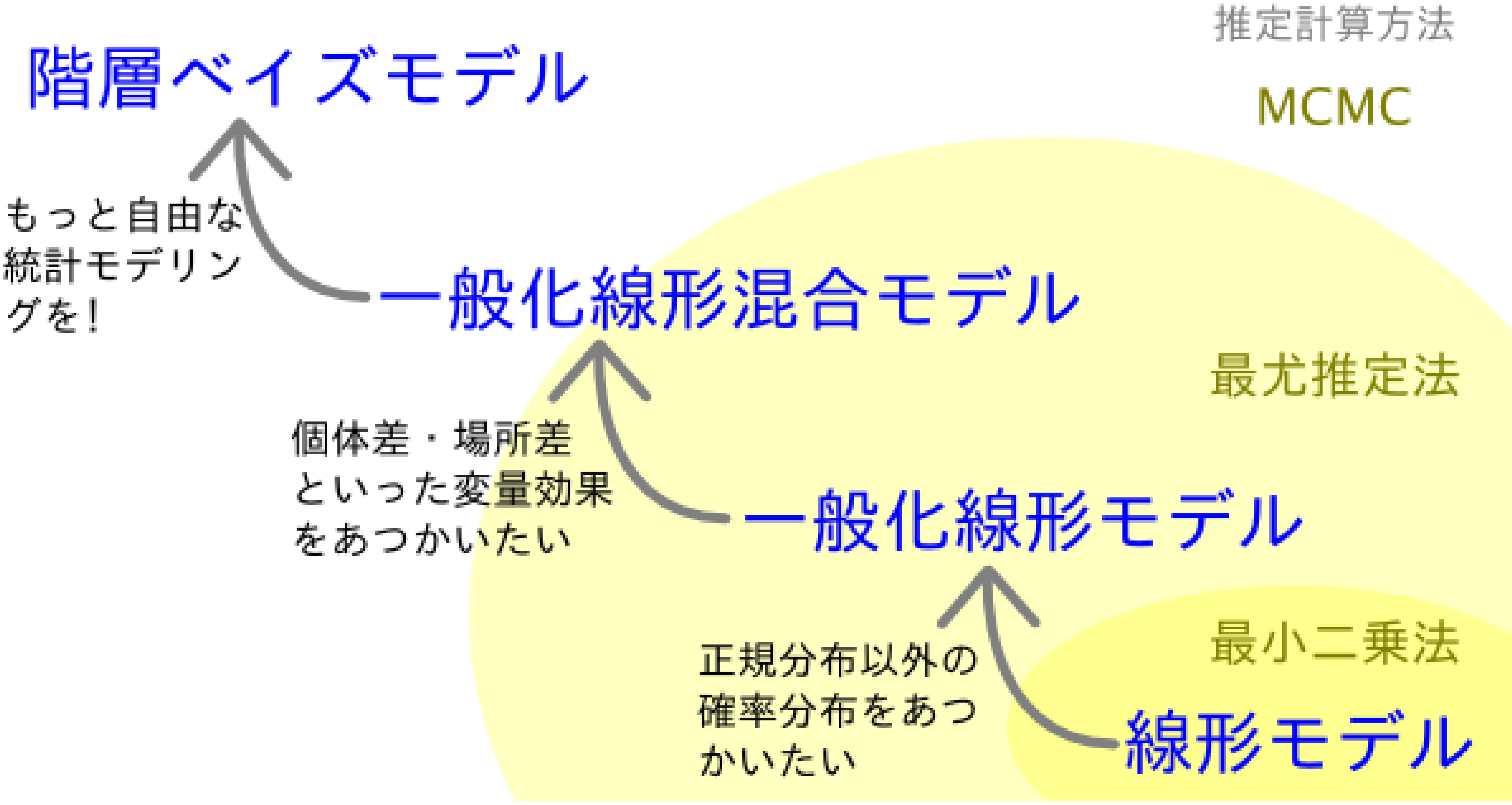
# 今回は尺が短いので

- とにかく、ざっくりと説明して、こんな方法もあるよねと言うことを確認
- 数学的な導出は省きまくります
  - (数式が好きな変態さんにはごめんなさい)
  - ふぇぇ;;(∩´﹏`∩);:
  - だって、行列がどうとか、ベクトルがどうとか、線形性がうんぬんかんぬんゆーても皆さん嫌でしょ？

# どうしてモデリング？

- 検定のような「差が有る」ことを示すだけでなく、データ全体の構造を知りたい
  - 検定だけでは分からない
- よくわかんない割り算や変数変換から脱出したい
  - そこまでして有意差にこだわるよりかは、モデリングと言う手段を考えてもよいのでは？

# 線形モデルの発展




http://hosho.ees.hokudai.ac.jp/~kubo/ce/LinksGlm.html

# 線形モデルって何なのさ？

- (一般)線形モデル :(general) liner model

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i + \varepsilon$$

- 単回帰分析(i = 1)
- 重回帰分析(i ≤ 2)
- t検定
- ANOVAやANCOVA
- もこのモデルで表すことが出来る

※こう表すこともできる

$$y = X\beta + e$$

# 単回帰分析

- $y_i = \alpha + \beta x_i + e_i$
- 書き換えると、

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

つまりは、

$$\longrightarrow \quad y = X\beta + e$$

# 重回帰分析(e.g., 2変量)

- $y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$
- 書き換えると、

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{n1} & x_{n2} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

つまりは、

$$\longrightarrow \quad y = X\beta + e$$

# ANOVA(e.g., 一元配置)

- 3つの水準を設定し、完全無作為法でそれぞれの処理を2回ずつ反復

$$\mathrm{y}_{\mathrm{ij}} = \mu + T_i + e_{ij}$$

- 書き換えると、

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ T_1 \\ T_2 \\ T_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix} \longrightarrow y = X\beta + e$$

つまりは、

# 一般線形モデルの特徴

- 簡単に言うと、応答変数yが<u>正規分布に従う</u>
- 詳しく言うと、<u>誤差(データのばらつき)が等分散正規分布</u>であることを仮定
  - この仮定を見たさない場合はどうしよう？
  - カテゴリカル、カウントデータの場合は仮定を満たさないので不適切？(∵離散変数だから)

※他にも特徴はありますが今回は割愛

# これでいいの？

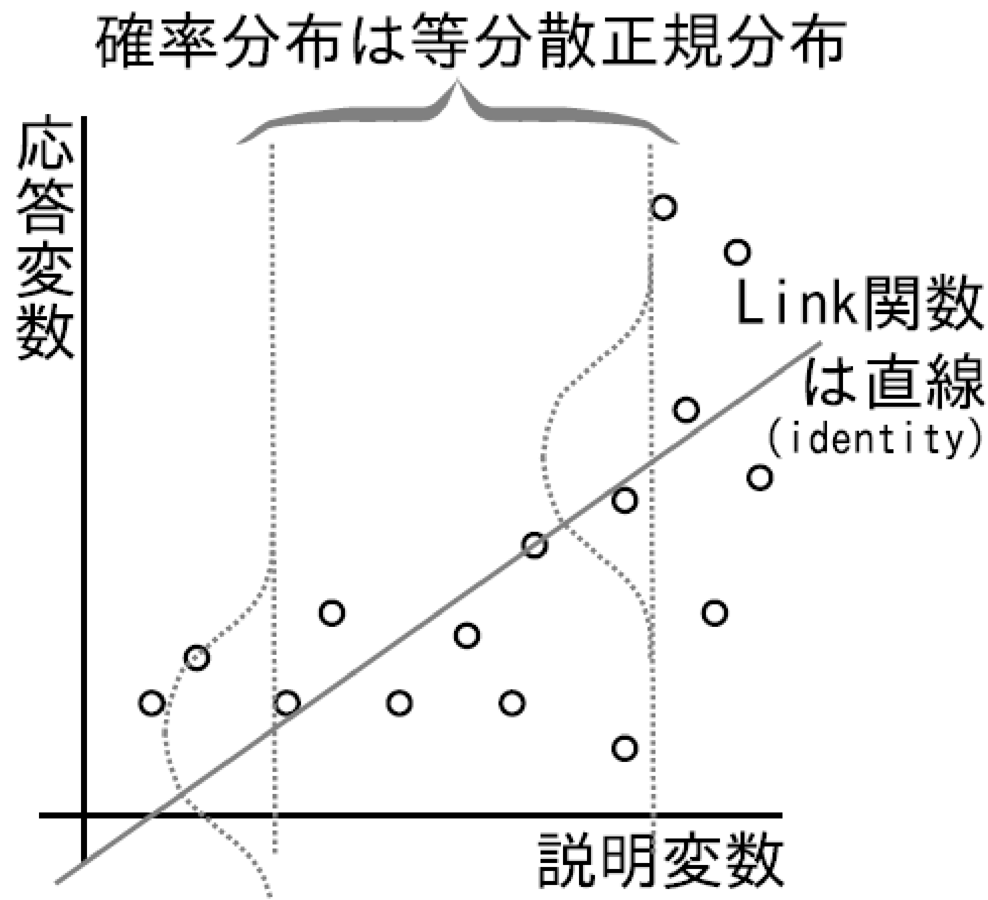



図 2: 古典的な直線あてはめ (一般線形モデル)

http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/49477/4/kubostat2008c.pdf

# こっちの方がそれっぽくない？

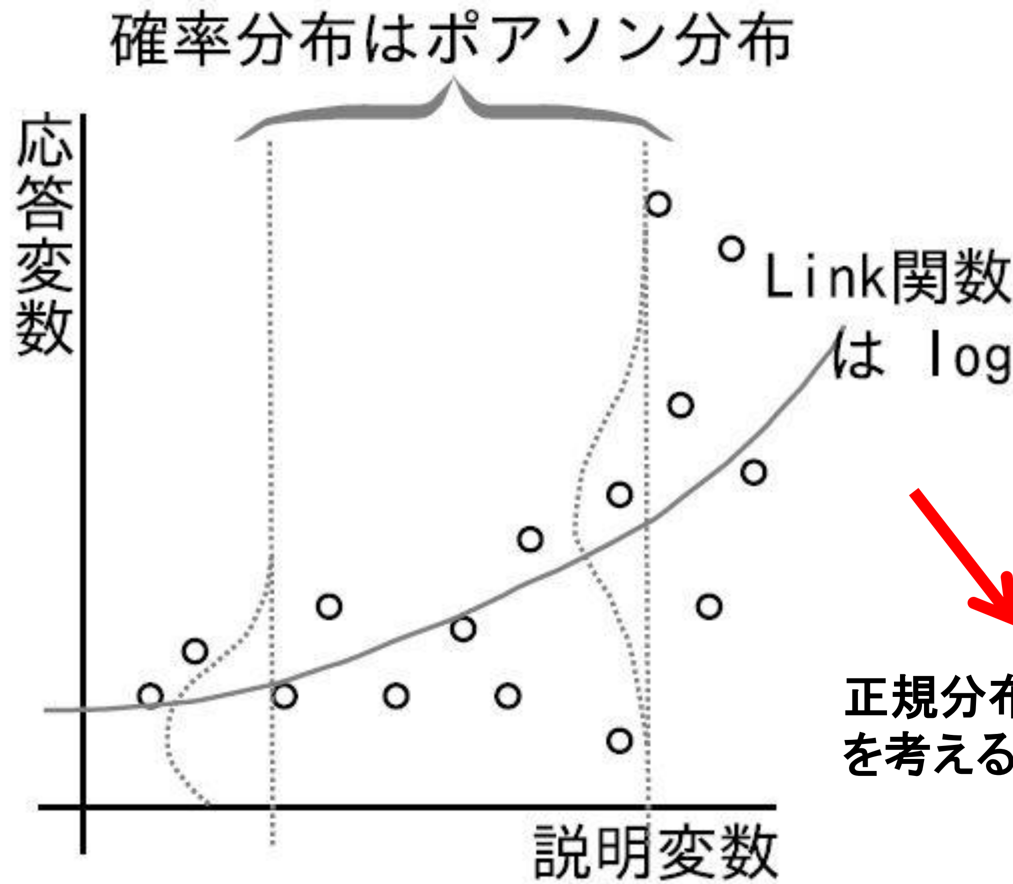



図 3: 一般化線形モデルによるあてはめの例

**正規分布以外への拡張**
**を考える必要がある！**

http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/49477/4/kubostat2008c.pdf

# そこで、一般化線形モデル (generalized liner model; GLM)の出番

- 正規分布だけでなく、指数型分布族(exponential family)と呼ばれる種類の分布も扱うことが出来る
  - e.g., 二項分布, ガンマ分布, ポアソン分布
  - これらはよい性質を持っている
  - 最尤推定しやすい、共役事前分布がある…

※他にも特徴はありま(ry

# GLMはどんなものなの？

- GLMを作るのに必要なパーツは次の3つ

① 確率分布

② link関数

③ 線形予測子(今回は省略)

  – 極端に言うと、確率分布が決まればだいたい大丈夫

# 確率変数

- 応答変数yがどのような確率分布に従うと考えられるのか？
- そのばらつきを正規分布だけでなく、二項分布やガンマ分布…を指定できる

# Link関数

- 式を変換して線形にする関数
- 分布によってlink関数はだいたい決まっている

| | 確率分布 | glm() の family | よく使う link 関数 | 分散 ($m$ は平均) |
|---|---|---|---|---|
| (離散) | ベルヌーイ分布 | binomial | logit | $\mu(1-\mu)$ |
| | 二項分布 | binomial | logit | $\mu(1-\mu)$ (注) |
| | ポアソン分布 | poisson | log | $\mu$ |
| (連続) | ガンマ分布 | gamma | log かな? | $\mu^2$ |
| | 正規分布 | gaussian | identity | 一定 |

http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/49477/4/kubostat2008c.pdf

# GLM with R

- 数学的導出は抜きにして、実際にやるにはどうすればいいの？
- Rでは単/重回帰分析とほとんど同じような感じでできます
  - SPSSさんは知りません…

# GLM with R

- e.g.,
- 応答変数が、ポアソン分布に従いそう
- とりあえず単回帰

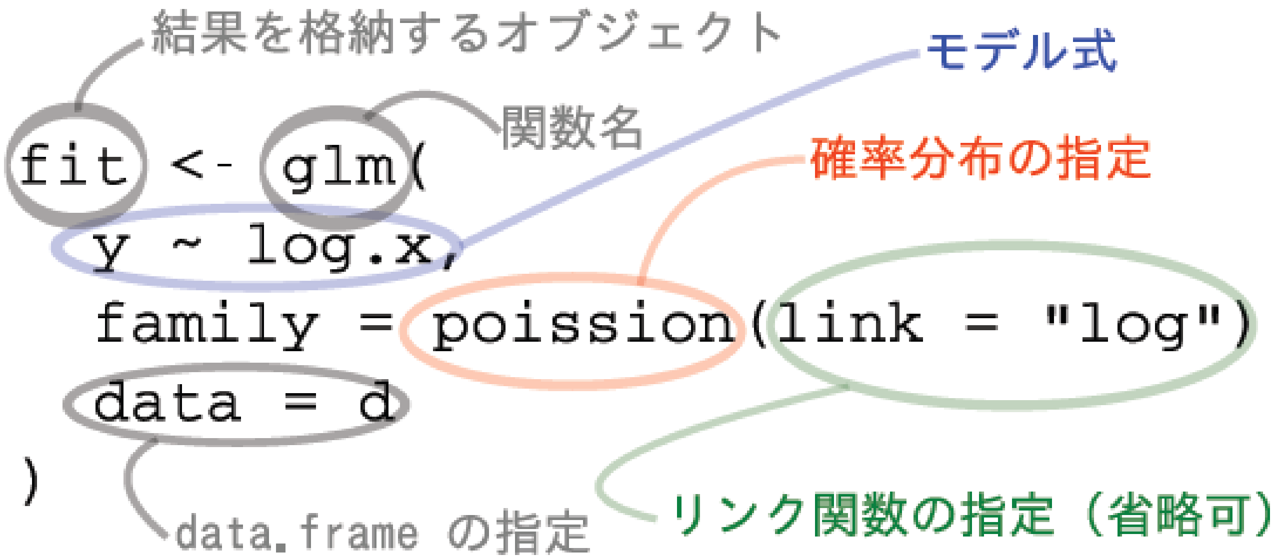
- glm(y ~ x, family = poisson(link = log),data = XXX)
- lm(y ~ x, data = XXX)

# GLM with R

- e.g.,
- 応答変数が、ポアソン分布に従いそう
- とりあえず単回帰

- glm(y ~ x, family = poisson(link = log),data = XXX)
- lm(y ~ x, ,data = XXX)

# GLM with R




- lmではなく、glmという関数を用いる
- モデル式：説明変数をどうするのか？
- 確率分布の指定：どんな確率分布にしたがうのか？
- link関数の指定

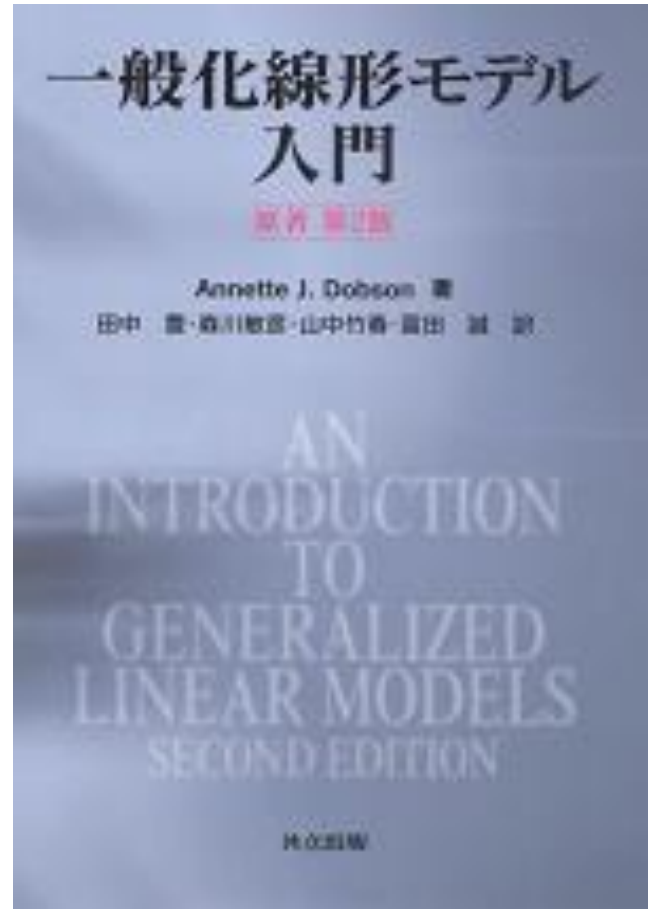
http:// http://hosho.ees.hokudai.ac.jp/~kubo/ce/2009/kubo2009glm.pdf

# まとめ

- 検定だけでなく、モデリング(モデル選択)という選択肢もあるよね
- GLMの構成要素：確率分布、link関数、線形予測子
  - 確率分布をうまく選ぶのが大事
- Rでやると、普通の回帰分析と同じくらいお手軽にできますよ

# 参考図書


確率と情報の科学
データ解析のための
統計モデリング入門
一般化線形モデル・階層ベイズモデル・MCMC
久保拓弥
岩波書店


一般化線形モデル
入門
Annette J. Dobson 著
AN
INTRODUCTION
TO
GENERALIZED
LINEAR MODELS
SECOND EDITION

# 引用文献


- 講義の一と : データ解析のための統計モデリング 第3回 http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/49477/4/kubostat2008c.pdf
- 生態学データの解析 ー GLM関連 http://hosho.ees.hokudai.ac.jp/~kubo/ce/LinksGlm.html
- 生態学のデータ解析 - 生態学会大会2009 http://hosho.ees.hokudai.ac.jp/~kubo/ce/2009/kubo2009glm.pdf

# 参考文献


- 北大 久保拓弥先生
http://hosho.ees.hokudai.ac.jp/~kubo/ce/FrontPage.html
- 土居正明さん
http://www012.upp.so-net.ne.jp/doi/
- 農環研 山村光司さん
http://cse.niaes.affrc.go.jp/yamamura/Intro.html
- 立教大 田中啓太さん
https://sites.google.com/site/keitaswebsite/connexions/osj-stat