

第7章

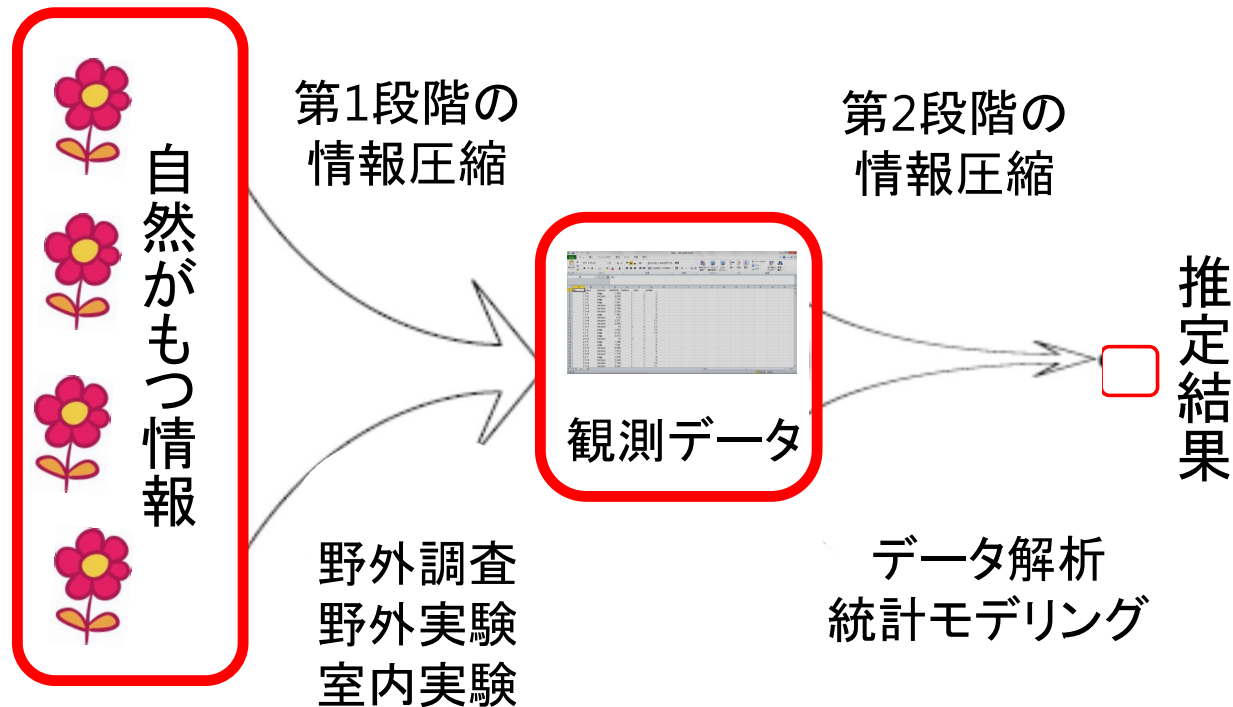
一般化線形混合モデル (GLMM)

一個体差のモデリング

本章の内容

実際のデータ解析でよく遭遇する「GLMではうまく説明できない」現象をうまくあつかえるようにGLMを強化する

◎観察者が測定できない・測定しなかった、個体や場所に由来する原因不明の差異＝「個体差・場所差」

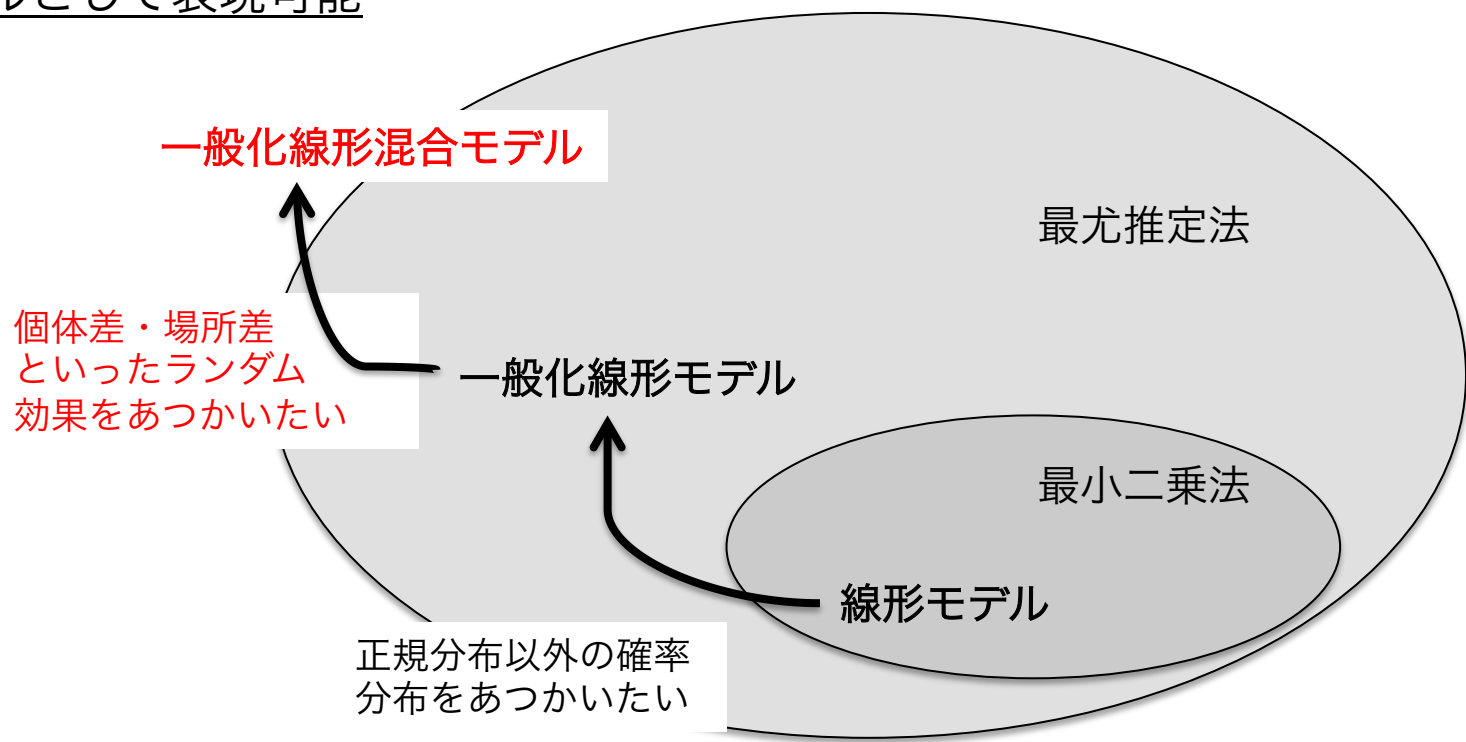


本章の内容

実際のデータ解析でよく遭遇する「GLMではうまく説明できない」現象を
うまくあつかえるようにGLMを強化する

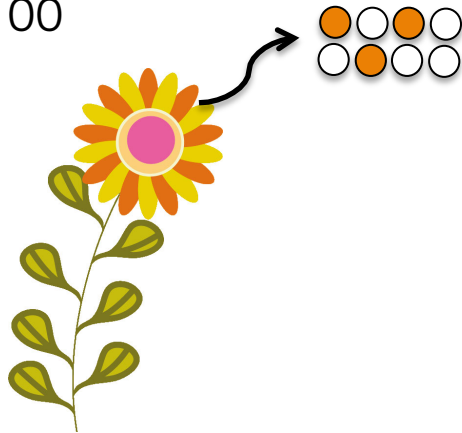
観察者が測定できない・測定しなかった、個体や場所に由来する原因不明の
差異＝「個体差・場所差」

→定量化はできないが、「何か原因不明がある」ことは
統計モデルとして表現可能



例題：架空植物の種子生存率

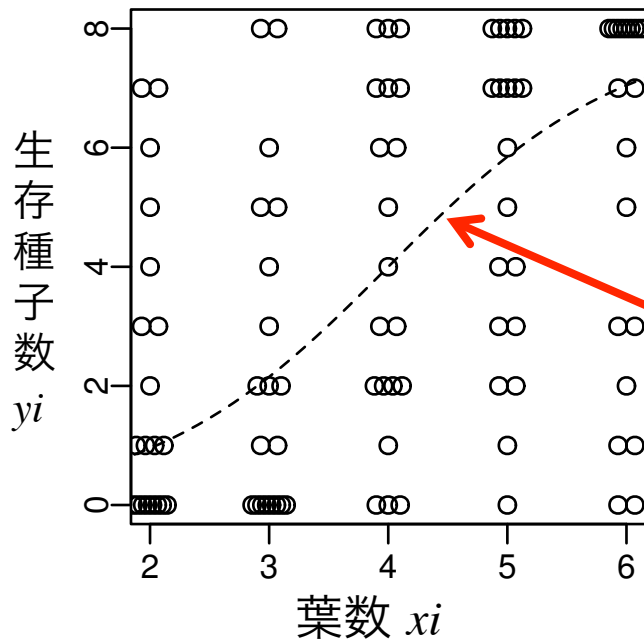
×100



調査個体数：100個体

各個体 i ごとの調査種子数 $N_i = 8$ 個

左の例の場合、生存種子数 $y_i = 3$

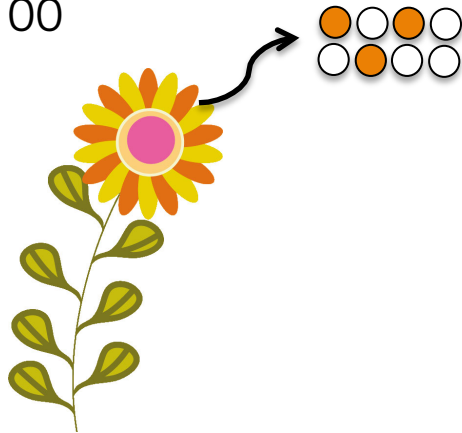


葉数 x_i が{2・3・4・5・6}であるものを
20個体ずつサンプリング

「真の」生存確率の一例

例題：架空植物の種子生存率

×100



種子の生存確率をGLMで推定してみる

$$\text{logit}(q_i) = \beta_1 + \beta_2 x_i$$

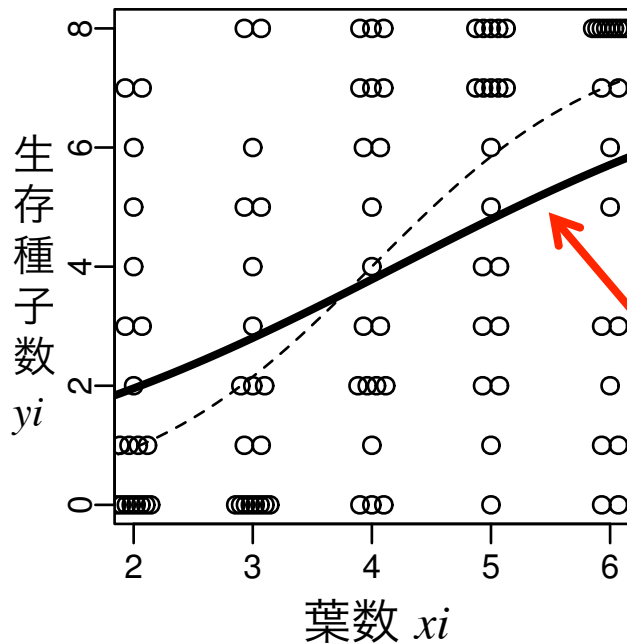
観測された生存種子数が y_i である確率が二項分布にしたがうとすると、

$$p(y_i | \beta_1, \beta_2) = \binom{8}{y_i} q_i^{y_i} (1 - q_i)^{8 - y_i}$$

全個体の対数尤度

$$\log L = \sum_i \log p(y_i | \beta_1, \beta_2)$$

$\log L$ が最大になる切片 β_1 と傾き β_2 をさがす
(最尤推定)



$$\begin{cases} \beta_1 = -2.15 \\ \beta_2 = 0.51 \quad (\text{真の傾き } \beta_2 = 1) \end{cases}$$

例題：架空植物の種子生存率

種子の生存確率をGLMで推定してみる

Rでやってみよう！

```
model <- glm ( cbind (y, N - y) ~ x, data = d, family =binomial)
```

summary(model)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.1487	0.2372	-9.057	<2e-16 ***
x	0.5104	0.0556	9.179	<2e-16 ***

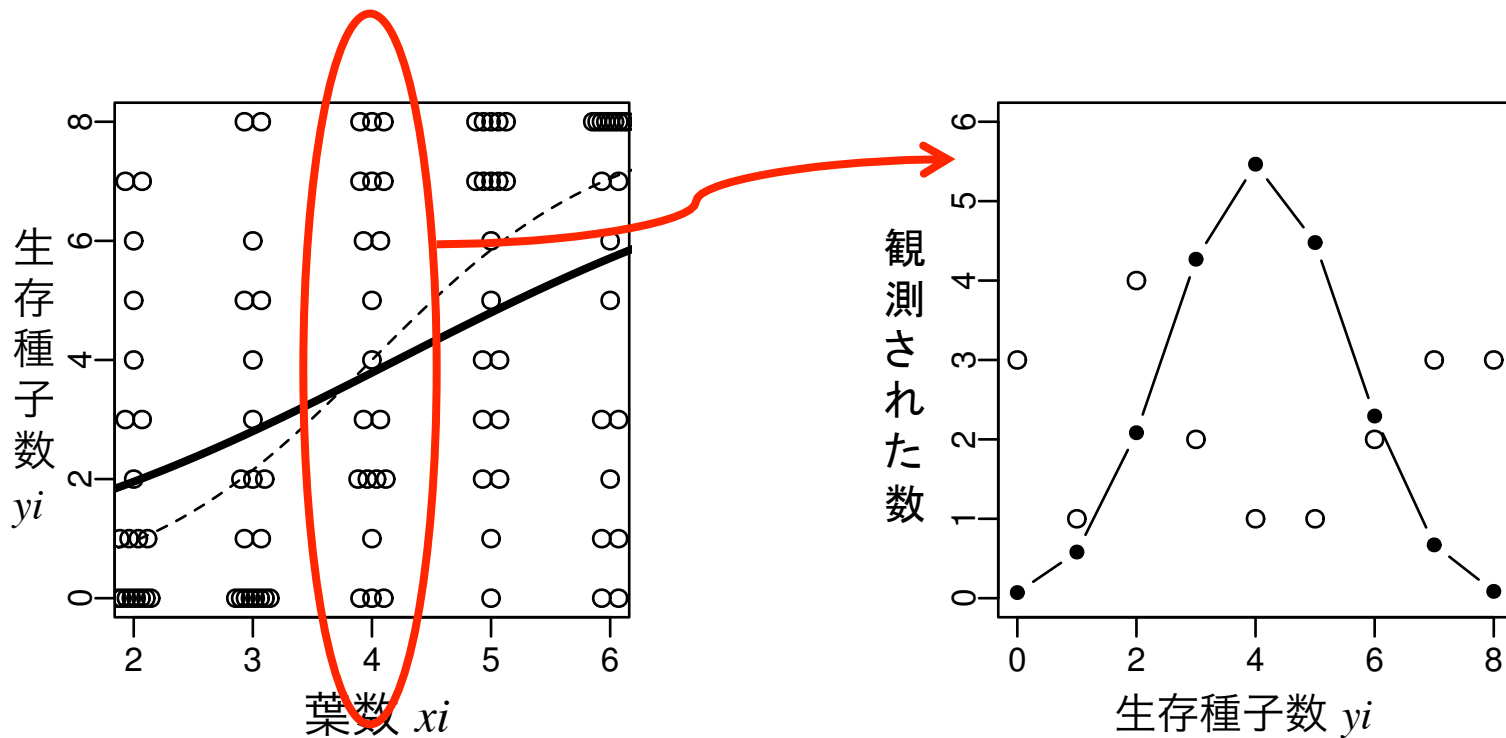
Null deviance: 607.42 on 99 degrees of freedom

Residual deviance: 513.84 on 98 degrees of freedom

AIC: 649.61

葉数 x_i は種子の生存確率に有意な影響を及ぼしている…と結論してOK??

例題：架空植物の種子生存率



$x_i = 4$ のときの

推定されたGLMから予測される二項分布（●と実線）と
観測された種子数分布（○）

→ ぜんぜん推定できていない！

過分散 (overdispersion)

統計モデルの仮定から期待されるばらつきよりも、実際のデータから計算されたばらつきの方が大きい状態

例題 ($x_i = 4$ のとき) で確認してみよう！

データフレーム d から葉数が $x_i = 4$ をみたすデータのサブセット $d4$ を抜き出す

```
d4 <- d [ d $ x ==4, ]
```

葉数が $x_i = 4$ であった個体の生存種子数 y_i を並べてみると

```
table(d4$y)
```

```
0 1 2 3 4 5 6 7 8  
3 1 4 2 1 1 2 3 3
```

これらの平均と分散は

```
C(mean ( d4 $ y ), var(d4 $ y ))  
[1] 4.050000 8.365789
```

平均 $Nq = 4.05$ なので

生存確率の平均 $q = 4.05 / N$

$= 4.05 / 8 = 0.5$ ぐらい

生存種子数 y_i が二項分布にしたがうなら生存確率の分散は：

$Nq(1-q) = 8 \times 0.5 \times (1 - 0.5) = 2$ ぐらいになるはずだが...

過分散が起こる原因

個体差：データとしては定量化も識別もされていないが、
「各個体（観測の単位）の何かに起因しているように見える差」

個体差をもたらす原因：生物的(biotic)な要因（遺伝子、年齢、経験など）

非生物的(abiotic)な要因（栄養や水分、光環境など）
→「場所差」あるいは「ブロック差」

このような「個体差」に影響をあたえている定量・特定することは不可能

→個体差や場所差を原因不明のまま、その影響だけをうまくとりこんだ
統計モデルが必要