実験デザインを反映した分散分析モデルとその意味

琉球大学大学院 理工学研究科 海洋環境学専攻
玉井 玲子・酒井一彦
devil_ray_oka@yahoo.co.jp

Decision Letter

Dear Dr. #######,

Please find attached files with referees' comments on your paper.

Your analysis is pseudoreplicated.

I therefore must reject your paper. Although this will be disappointing news, I hope the referees' comments will help you to prepare your paper for another journal.

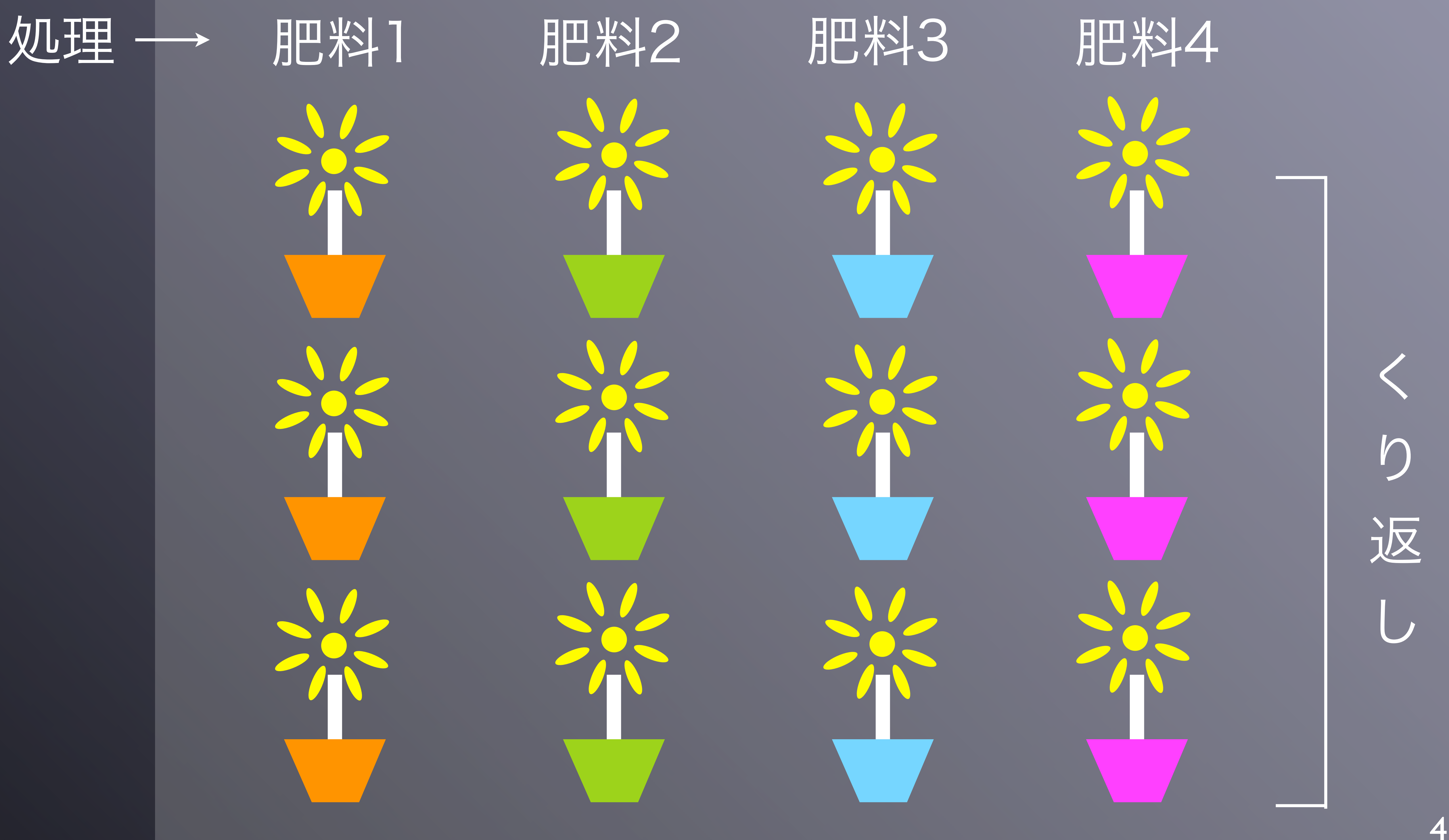
Yours sincerely,

Editor

I. 分散分析の数学的原理

一元配置分散分析の例

4種類の肥料で、植物の成長率を比較する



一元配置分散分析の例

一般化した結果の例

処理1	処理2	処理3	…	処理 i	…	処理 a	
y_{11}	y_{21}	y_{11}	…	y_{i1}	…	y_{a1}	
y_{12}	y_{22}	y_{12}	…	y_{i2}	…	y_{a2}	
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
y_{1j}	y_{2j}	y_{1j}	…	y_{ij}	…	y_{aj}	
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
y_{1n}	y_{2n}	y_{3n}	…	y_{in}	…	y_{an}	
$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$	$\bar{y}_{3\cdot}$	…	$\bar{y}_{i\cdot}$	…	$\bar{y}_{a\cdot}$	$\bar{y}_{\cdot\cdot}$

各群の平均

総平均

知りたいこと

各群の平均値は互いに異なると言えるか？

分散分析モデルとは

観測値をいくつかの要素の影響の和として捉える考え方

i 番目の処理の j 個目の観測値

$$y_{ij} = \mu + \alpha_i + e_{ij} \quad (i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, n)$$

全体の平均

i 番目の処理の効果

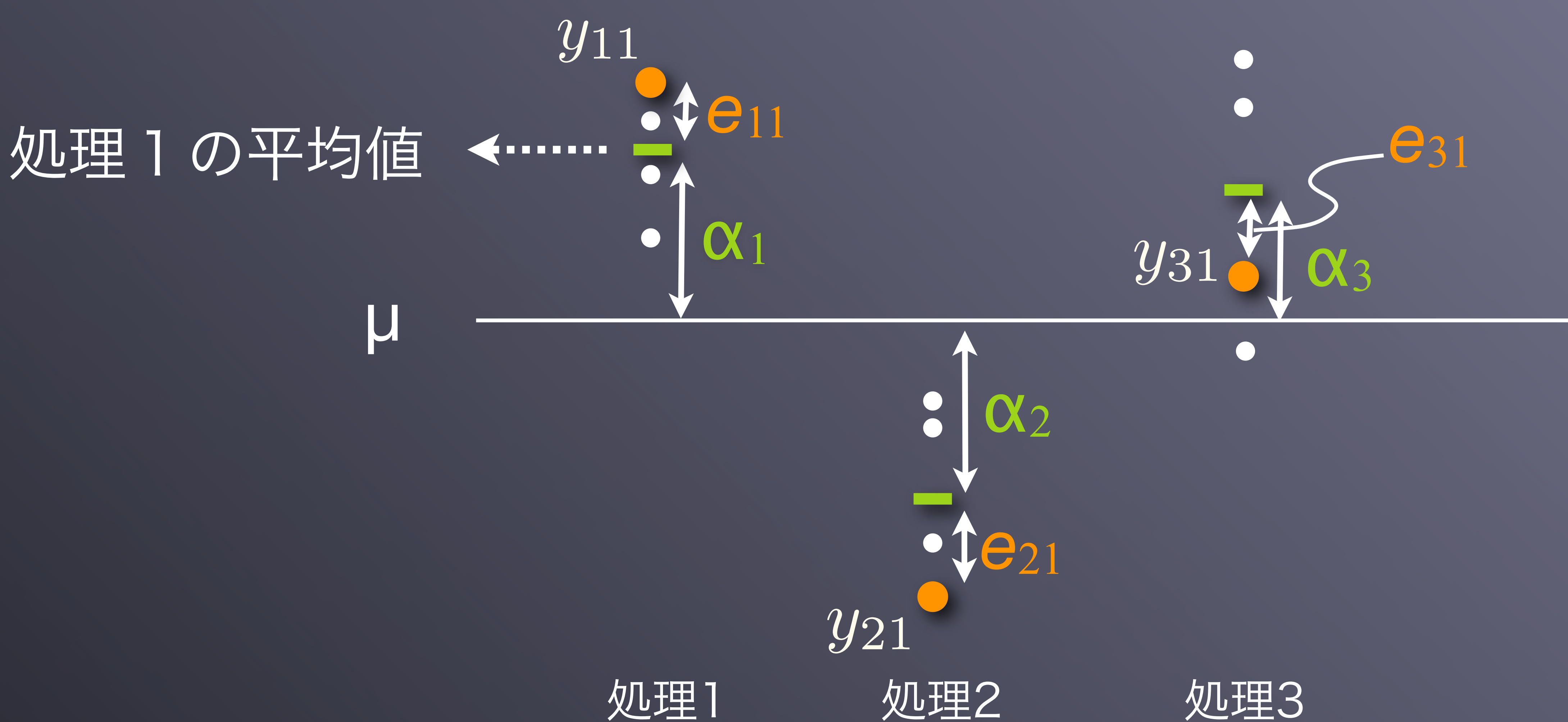
残差（誤差項）

||

数式には還元できない個々のデータのばらつき

モデルの各項のイメージ

$$y_{ij} = \mu + \alpha_i + e_{ij} \quad (i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, n)$$



帰無仮説：各群

$\Leftrightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$

F検定を用いて判定する

分散分析におけるF検定

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

分散分析におけるF検定

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

$$y_{ij} = \bar{y}_{\cdot\cdot} + (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot})$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2$$

SS_T	SS_B	SS_W
全体の平方和	群間の平方和	残差の平方和

基礎知識: χ^2分布とF分布

- χ^2分布

平均 μ, 分散 σ^2 の正規分布

N (μ, σ²) から、変数 X_i をn個ランダムサンプリングしたとき

$$\underbrace{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}} \sim \chi^2[n-1]$$

この値が **自由度 (n-1)** の χ^2分布に従う

- F分布

χ^2分布に従う2つの変数 V_1 と V_2 があるとき：

つまり $V_1 \sim \chi^2[v_1],\ V_2 \sim \chi^2[v_2]$

$$\underbrace{\frac{V_1/v_1}{V_2/v_2}} \sim F[v_1,\ v_2]$$

この値が **第1自由度 v_1, 第2自由度 v_2** のF分布に従う

分散分析におけるF検定

群間平方和 SS_B と 誤差平方和 SS_W を用いて F値を計算する。

$$F = \frac{\hat{\sigma}_1{}^2/\sigma_1^2}{\hat{\sigma}_2{}^2/\sigma_2^2}$$

← 主効果によるばらつき

← 誤差によるばらつき

$$F_{cal} = \frac{\left(\boxed{\frac{n}{a-1}\sum_{i=1}^{a}\alpha_i^2} + \hat{\sigma}_e^2\right)/\sigma_e^2}{\hat{\sigma}_e^2/\sigma_e^2}$$

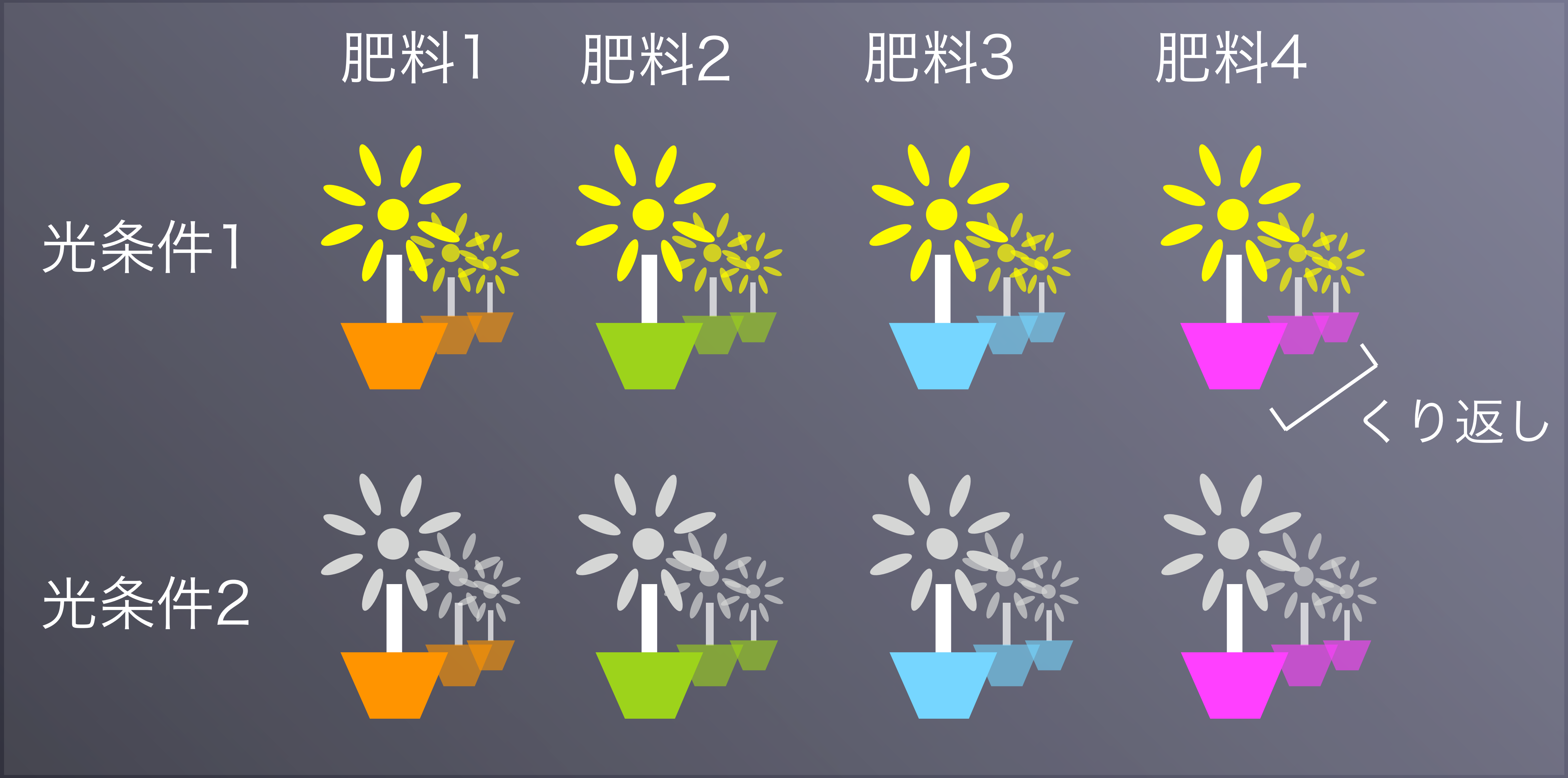
帰無仮説が正しいとき

$\alpha_1 = \alpha_2 = ... = \alpha_a = 0$

なので、0 になる

多元配置分散分析の考え方

4種類の肥料と2種類の光条件で植物の成長率を比較する



分散分析モデル：

要因Aと要因Bの交互作用

$$y_{ijk} = \mu + \alpha_i + \beta_j \qquad\qquad + e_{ijk}$$

$$(i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b,\ k = 1, 2, \ldots, n)$$

分散分析の数学的仮定

- 正規性
  - F分布自体、正規分布から派生したものであるため
- 等分散性（母分散が等しい）
  - 正規分布は平均と分散によって形が決まるため
  - もし等分散でなかったら、帰無仮説が棄却された理由が「平均値の違い」なのか「分散の違い」なのか区別できない
- 独立性
  - 観測値の残差の期待値が互いに無関係

II. 残差の独立性と要因の意味

II-1. 残差が独立でないことで生じる問題

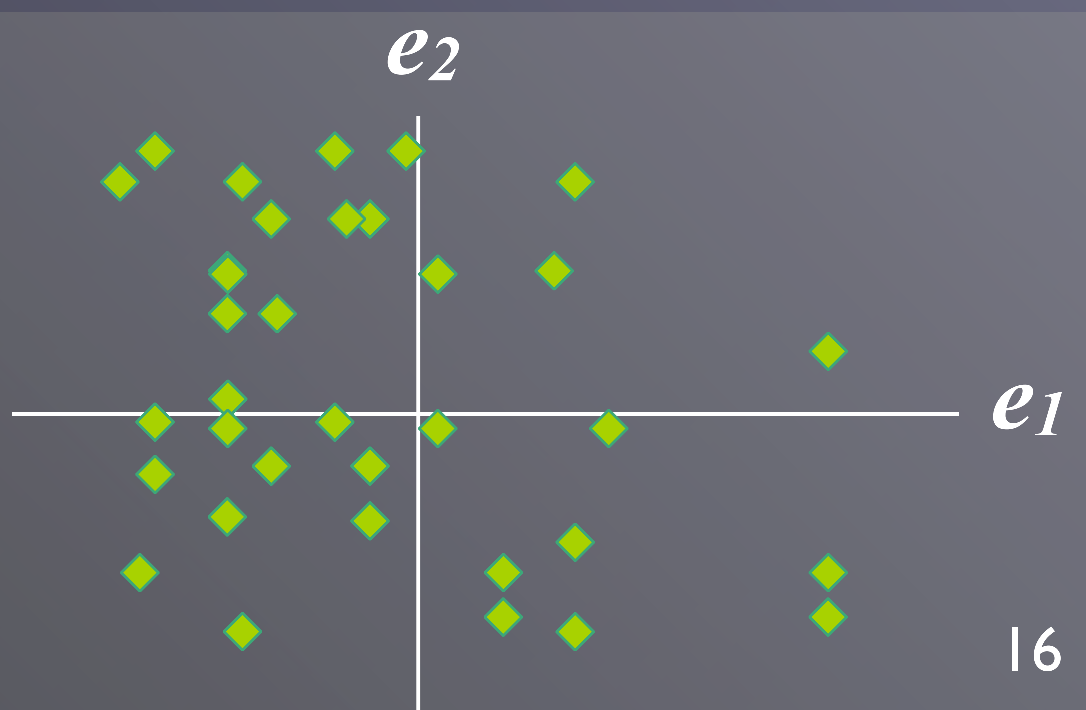
残差が独立でないことで生じる問題

「残差が独立でない」って何?

それの何がいけないの?

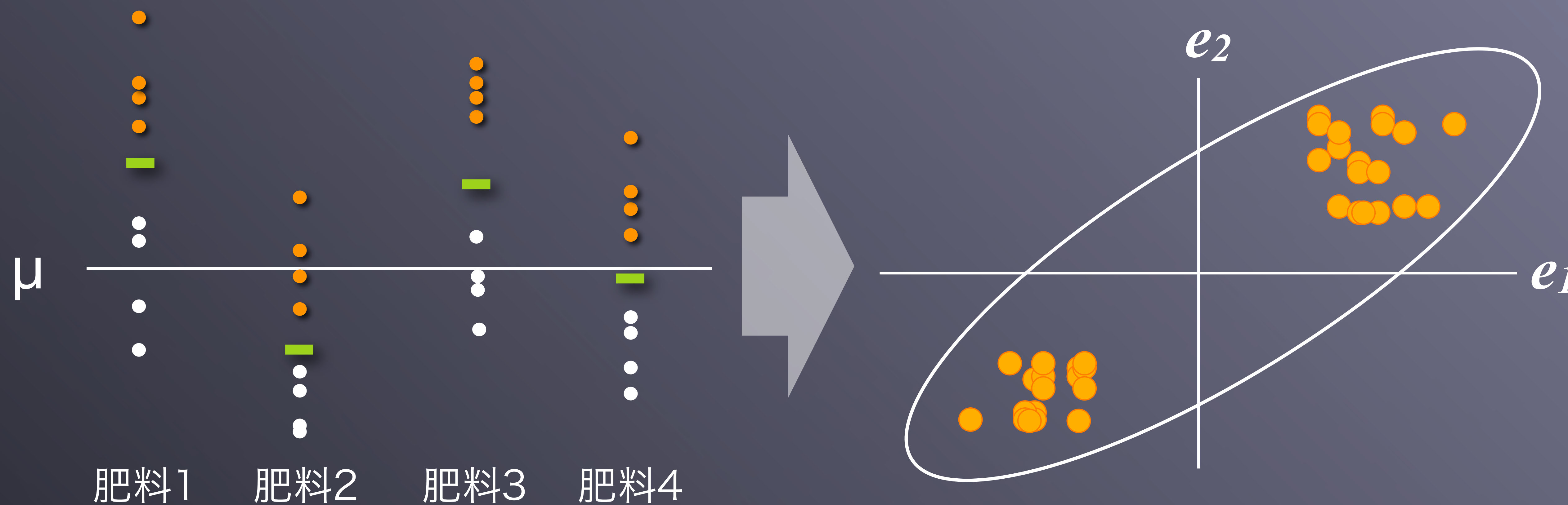
先ほどの二元配置の実験

同じ水準から取り出した
2つの観測値の残差を
それぞれ x, y に
プロット



必要な要因を入れなかった場合

- 光条件を無視して肥料の効果のみで解析

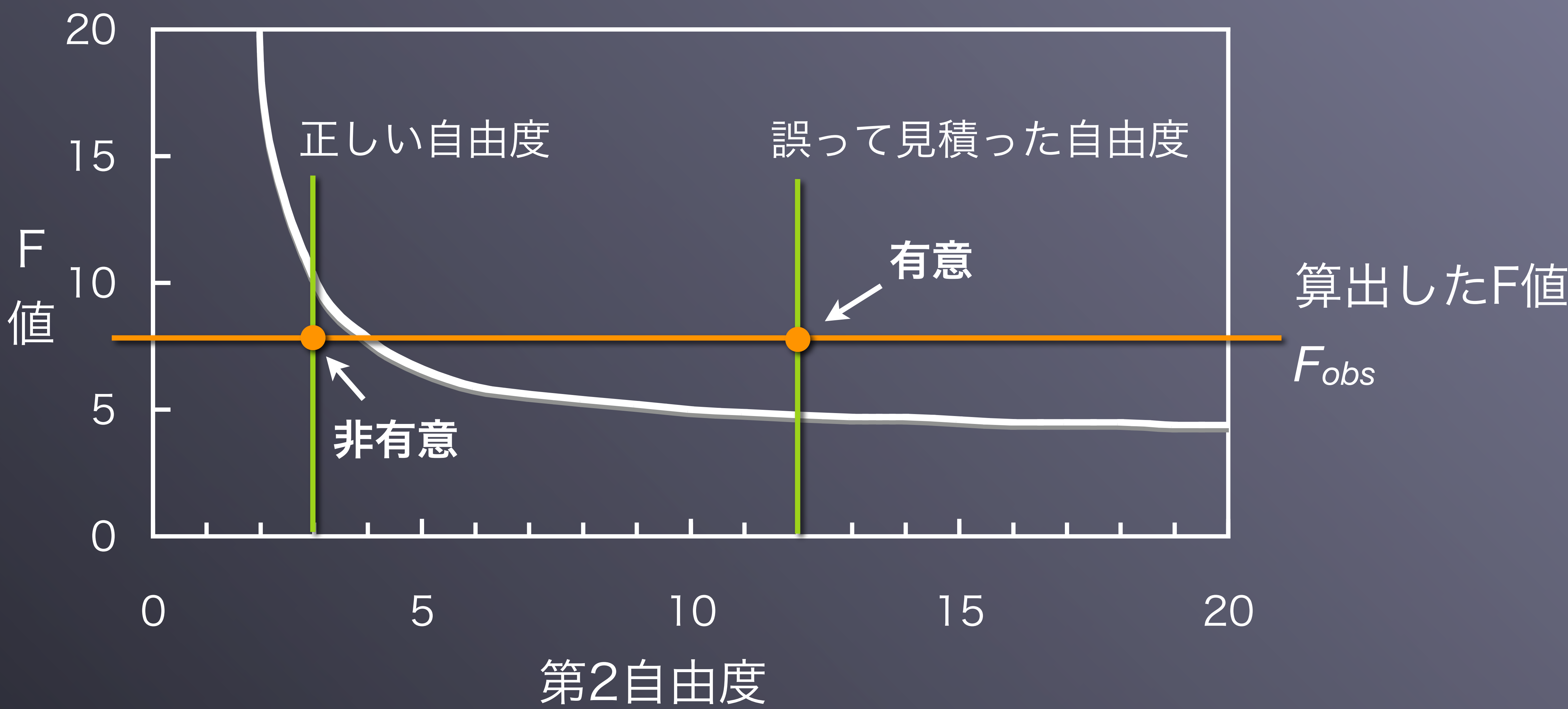


緑のバーから各点までの距離が残差

本来分けるべき群ごとに残差を2つずつ取り出してプロット

偽反復によって起こる問題

残差の相関が正 → 自由度の過大評価



第1自由度を固定、様々な第2自由度に対してp=0.05となるF値をプロット

帰無仮説を誤って棄却してしまう確率が増大

II. 残差の独立性と要因の意味

II-2. 実験単位と観察単位の不一致 (実験例1)

実験単位と観察単位

- **実験単位** experimental unit：

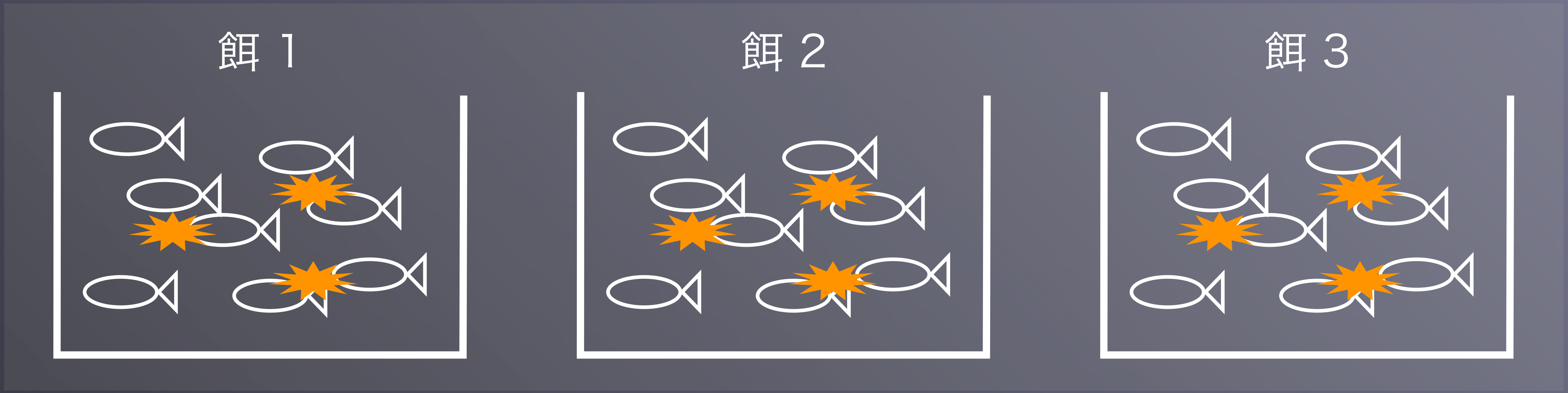
  統計的に独立な最小単位

- **観察単位** observational unit ：

  実際に観察を行う単位

実験単位と観察単位が一致しない例

餌の違いによる魚の成長率の違いを調べる実験



このデザインは一元配置？

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

要因：餌の種類
繰り返し：個体

水槽内で魚が餌を取り合う場合

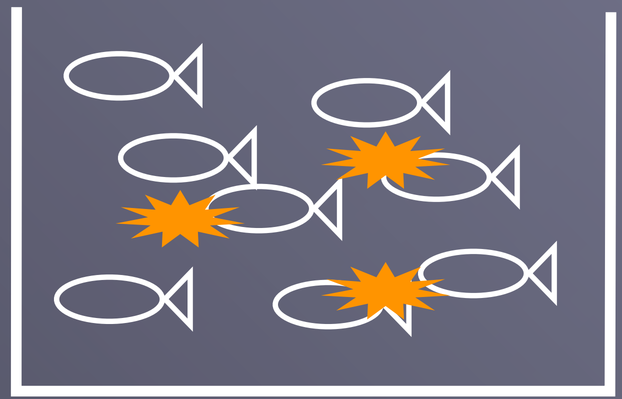
- 魚の成長率に対して餌の種類以外の要因がはたらく
- 個体間関係は足し算でモデルに組み込むことができない

実験単位と観察単位

- **実験単位** experimental unit：

  統計的に独立な最小単位

  ⇒ ex) 水槽



- **観察単位** observational unit ：

  実際に観察を行う単位。

  ⇒ ex) 個体

II. 残差の独立性と要因の意味

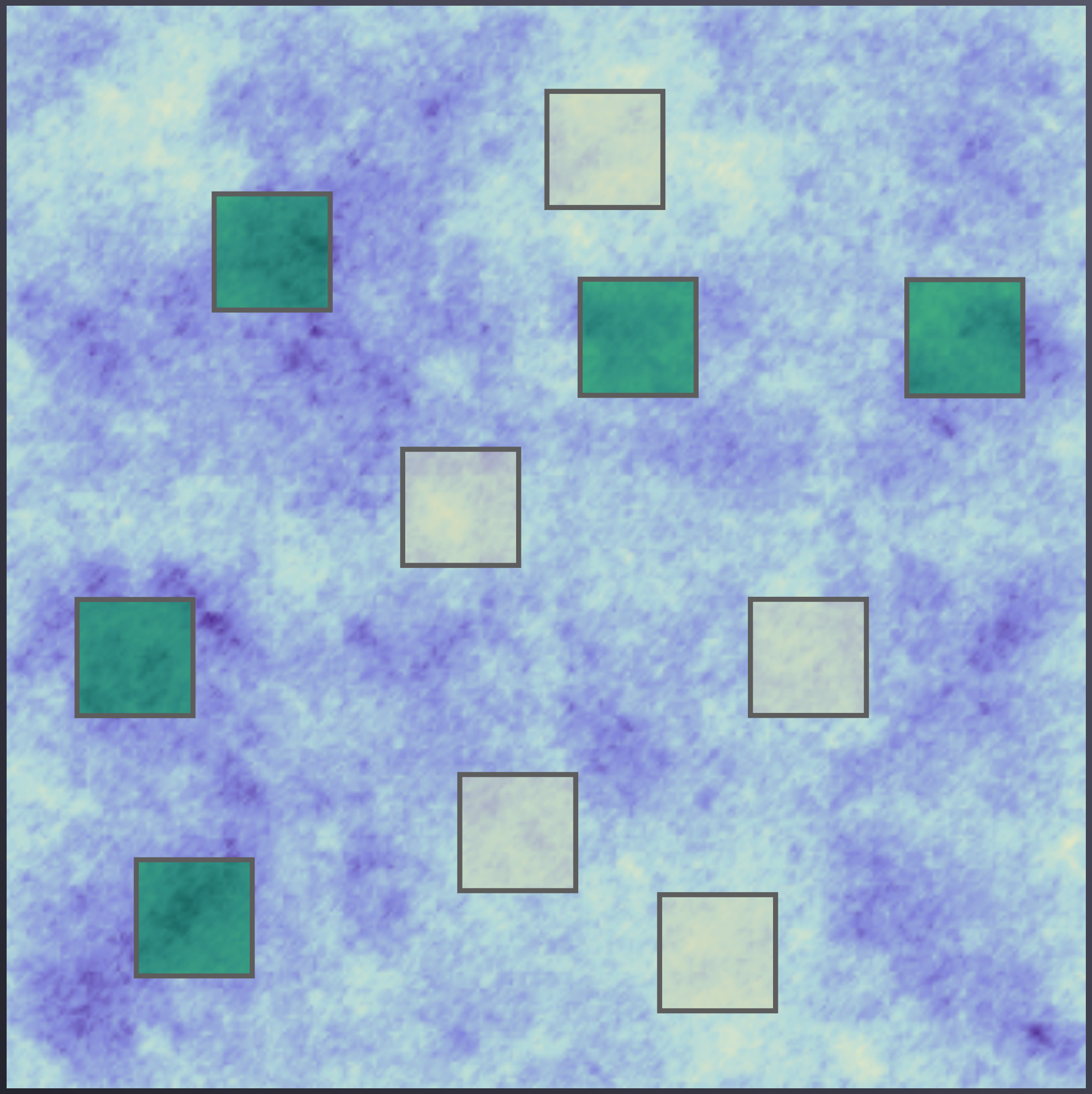
II-3. 空間的自己相関へのブロッキングによる対処 (実験例2)

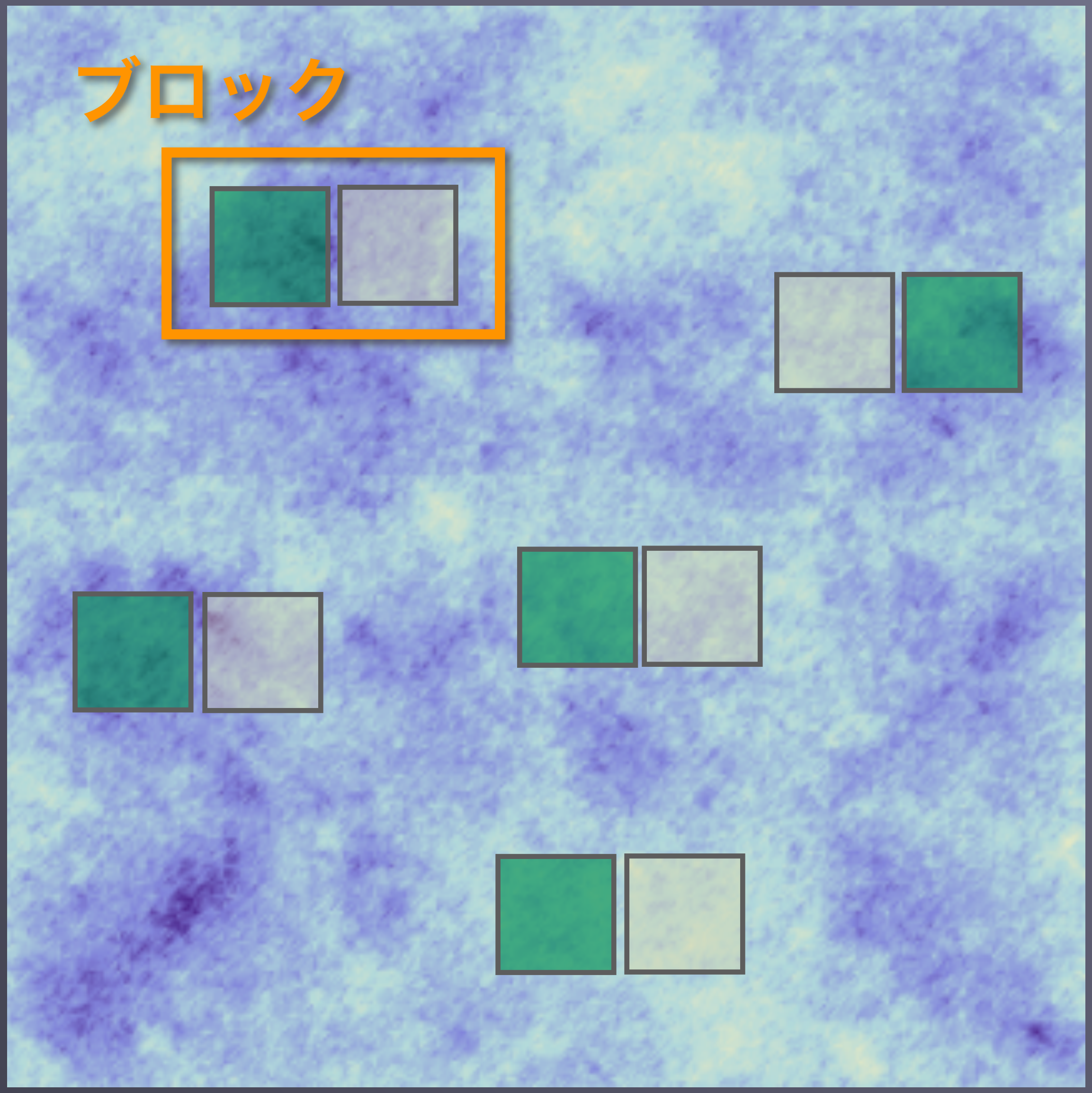
空間的自己相関 ：場所が近いと環境条件が似ること

野外実験…環境の不均一性を正確に把握できない

完全ランダム化デザイン

ランダマイズド ブロックデザイン

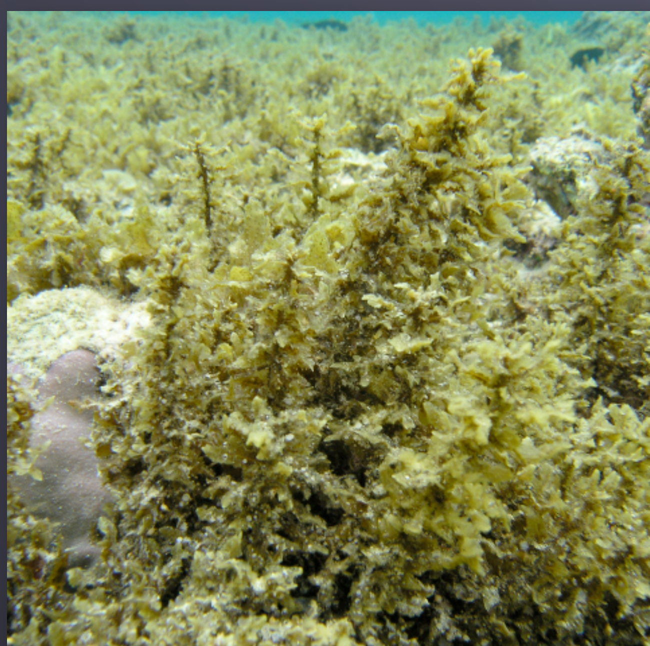




処理1　処理2

背景

空間競争



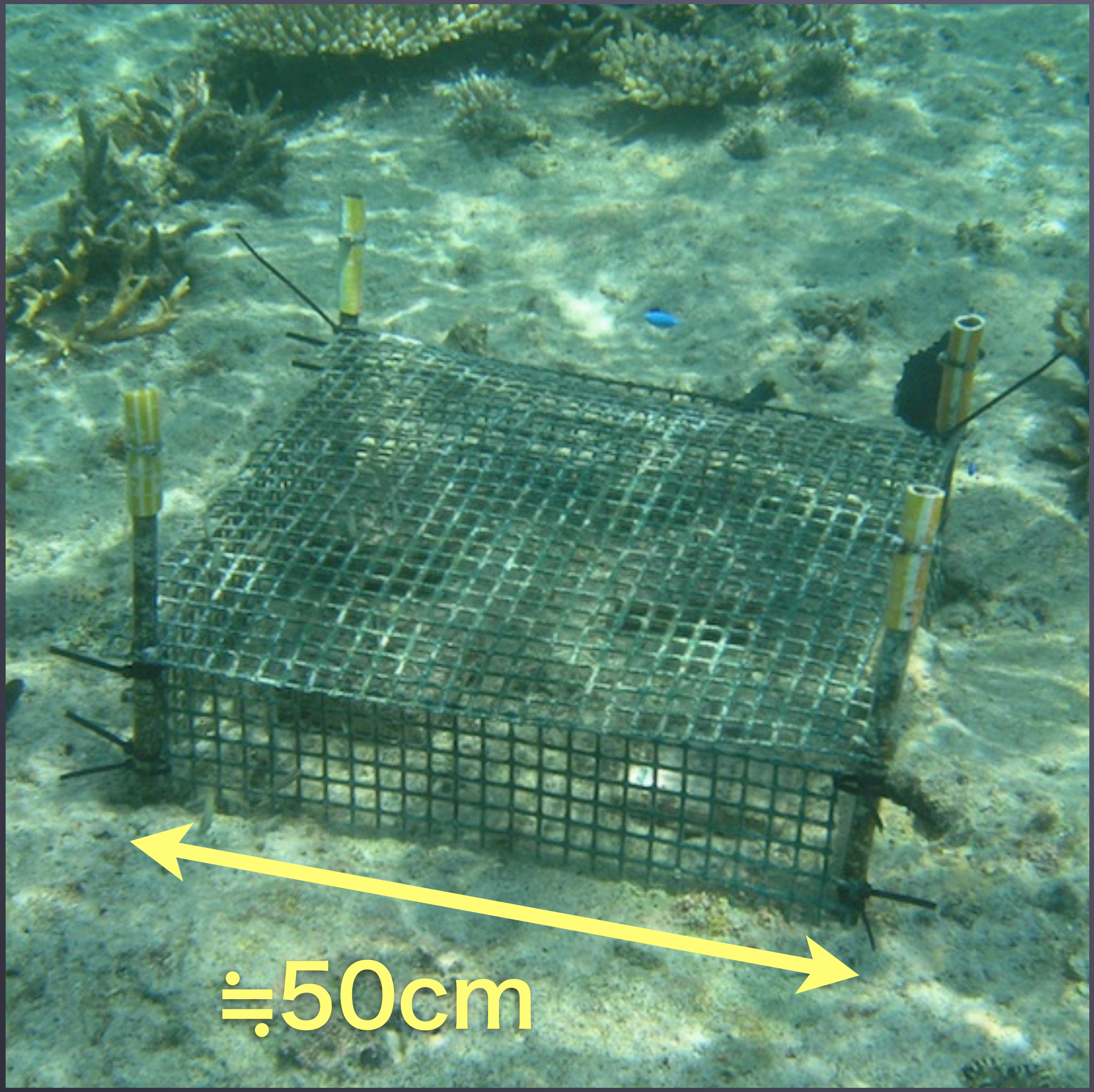
促進

富栄養化

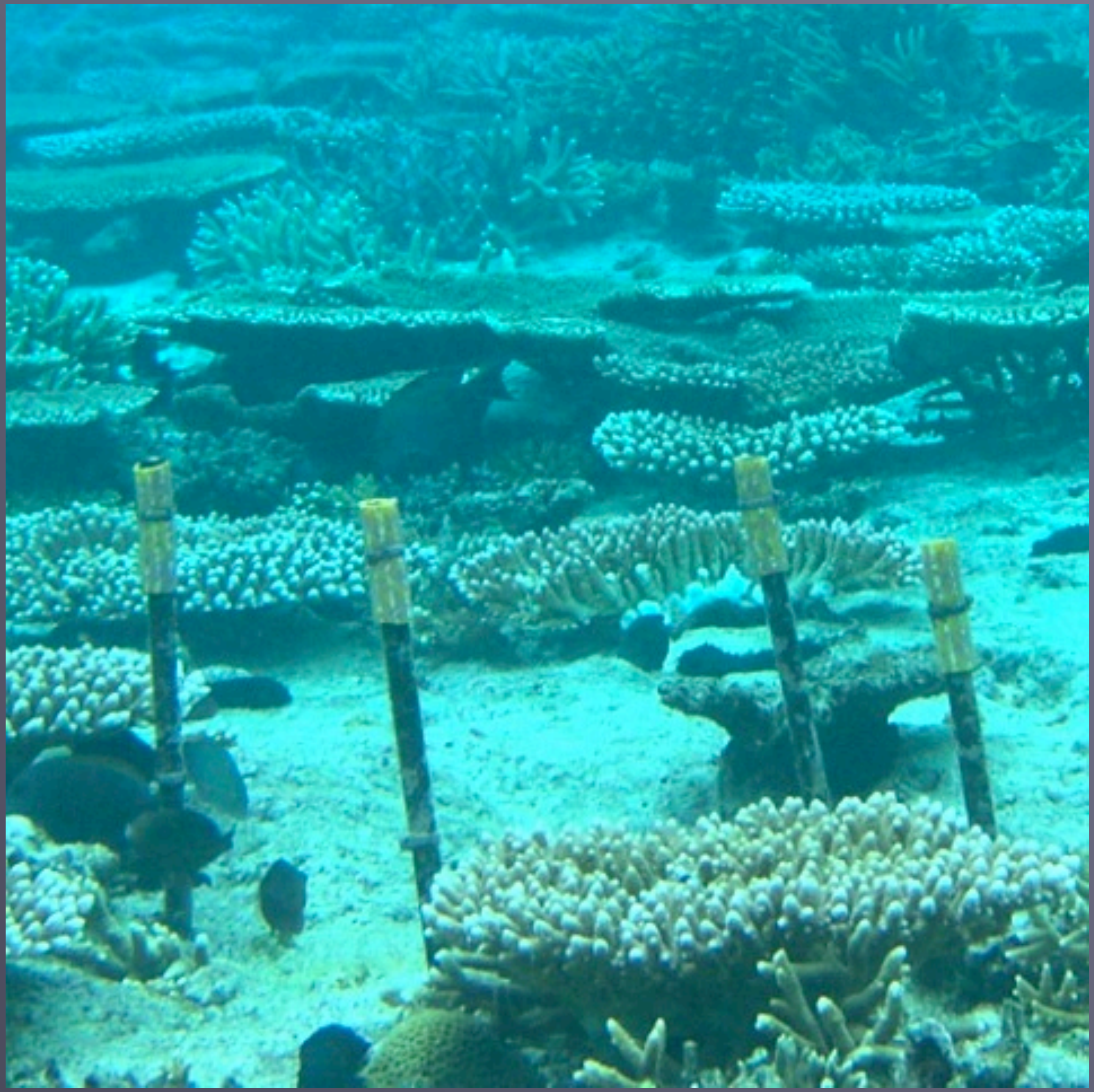
目的

藻食性動物の減少が海藻の増加を通して
サンゴの成長率に及ぼす影響を調べる

実験手法: ケージング



ケージあり



ケージなし

ケージで藻食性動物を排除

実験デザイン

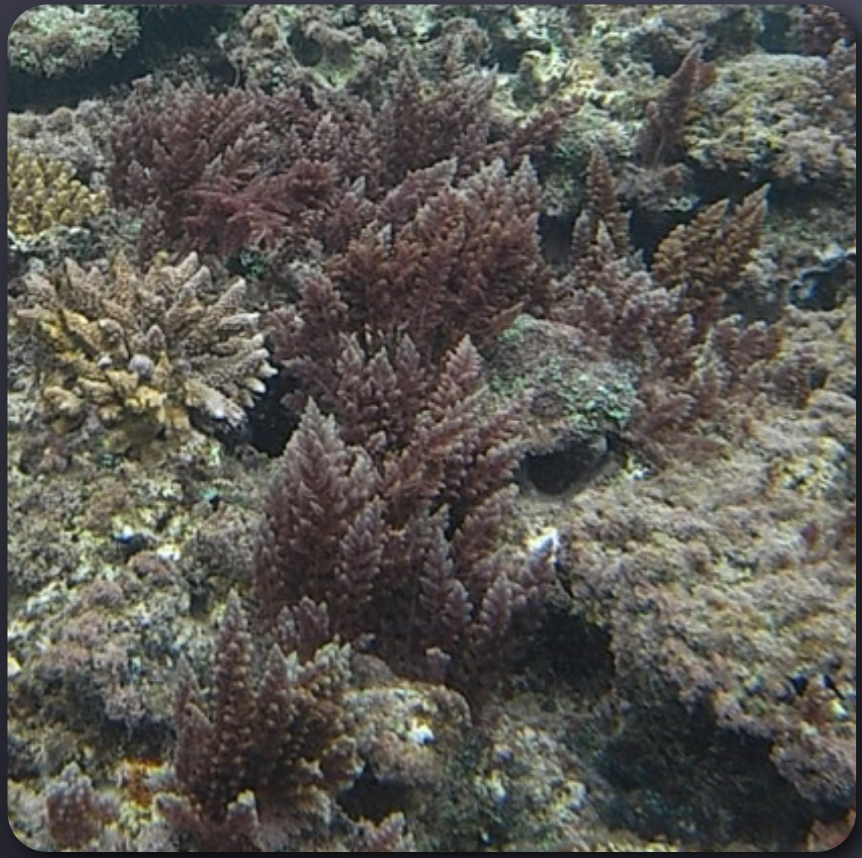
解析に含める要因

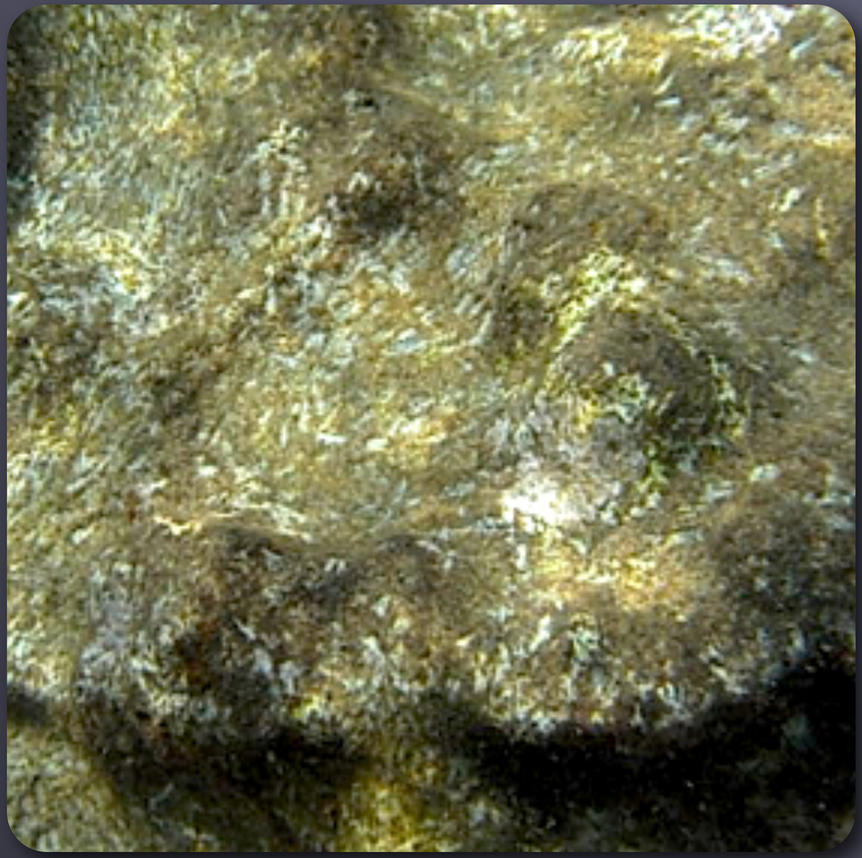
- 処理 …ケージあり or なし

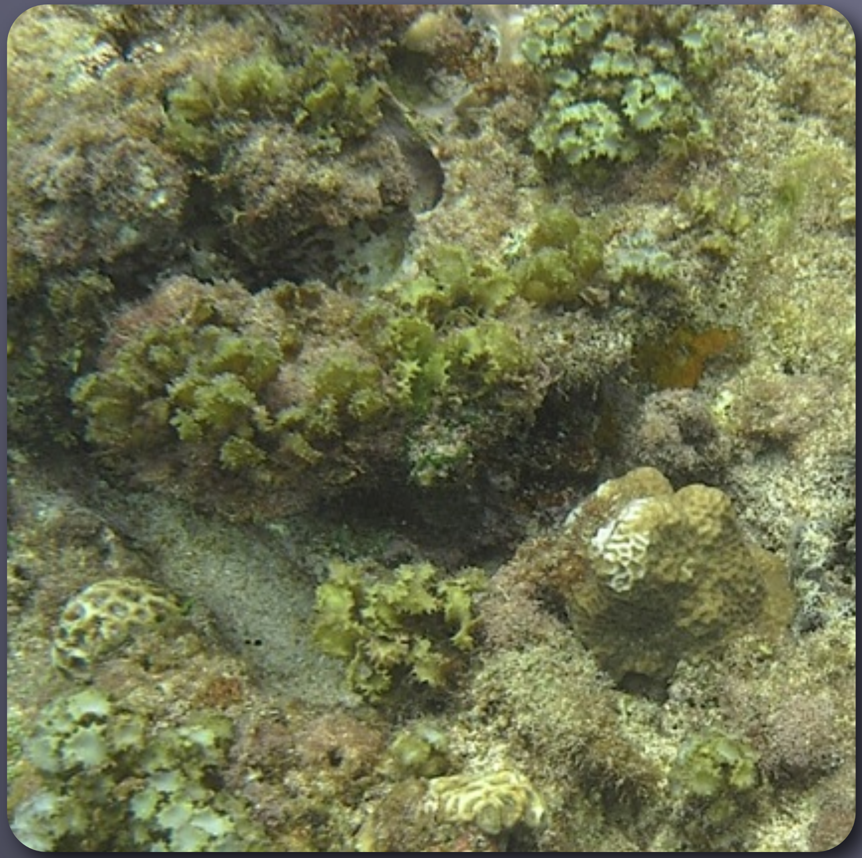
実験デザイン

解析に含める要因

- ブロック
- 処理 …ケージあり or なし



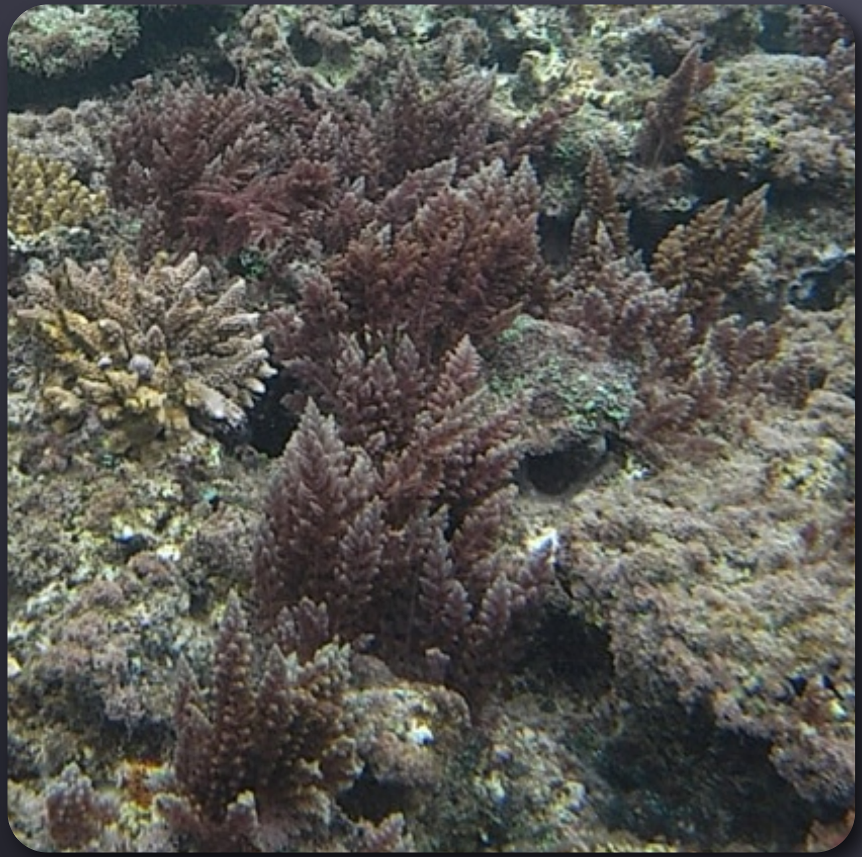


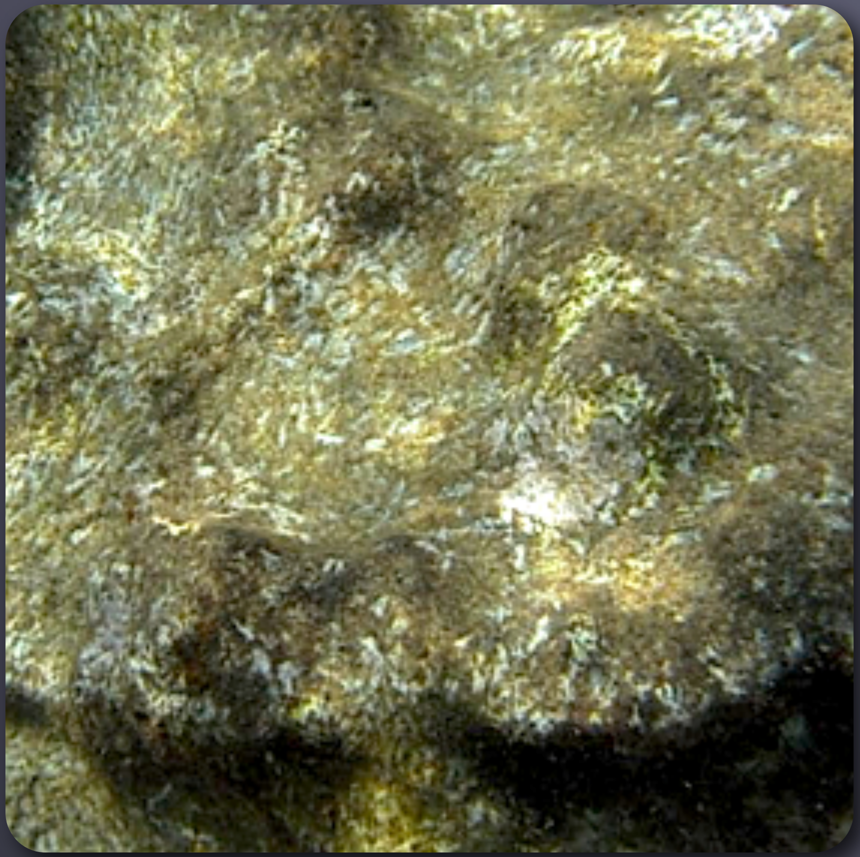


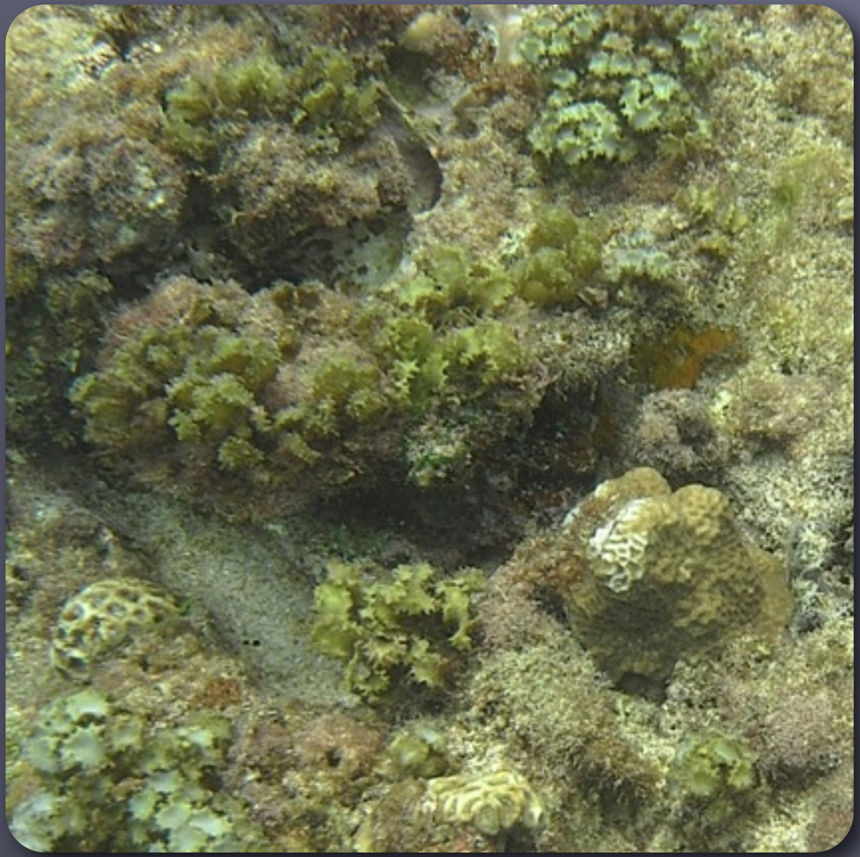
実験デザイン

解析に含める要因

- エリア
- ブロック[エリア]
- 処理 …ケージあり or なし







まとめ

I. 分散分析の基礎の確認

- 分散分析の数学的原理
- 分散分析の前提条件

II. 残差の独立性と要因の意味

- 残差が独立でないことで生じる問題点
- 実験単位と観察単位の不一致 (実験例1)
- 空間的自己相関へのブロッキングによる対処 (実験例2)

まとめ