# 分散分析

馬場　真哉
解析　U4

# 今日話す内容

３つ

- 分散分析とはなにか
- 分散分析の計算
- 教科書内容　（Ｒによる実演も）

# 分散分析とは何か？

# 分散分析とは何か

内容３つ

- 分散分析の使いどき
- 分散分析な考え方
- 分散と言う”威力”

# 分散分析の使いどき　１

## 「平均値に差があるか？」

が気になる人はとりあえずこれ。

- ３つ以上の水準云々とかは忘れて可
- t 検定とかも忘れても可

⇒ 分散分析したって結果は一緒だから

（等分散の条件付きだけど）

# 分散分析の使いどき　2

例

- 肥料は本当に収穫量を高めてくれているか？
- 水温 15 ，20 ，25 度で、イカの平均摂食量に差があるか？
- 使用した薬 A,B,C の間で魚の体長の平均値が変化するか？

⇒ この例を使用する

# 分散分析の考え方　１

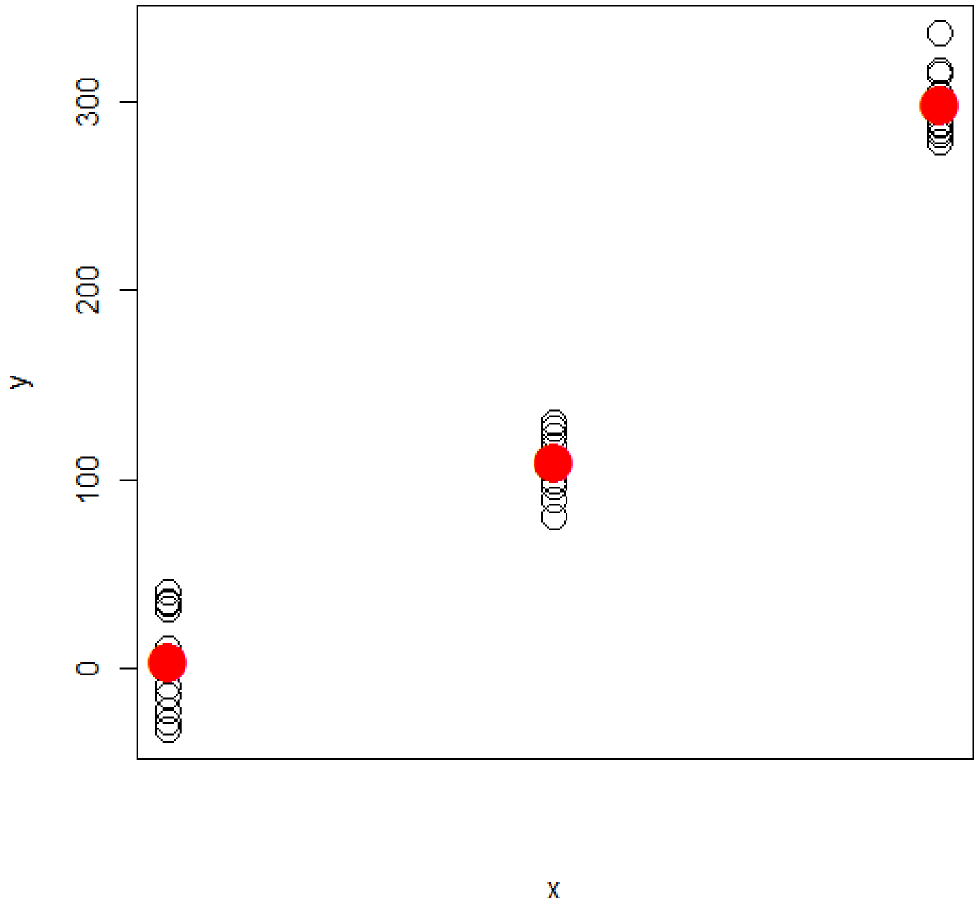
## 変動の分解

## ⇒ 誤差と効果を分離する

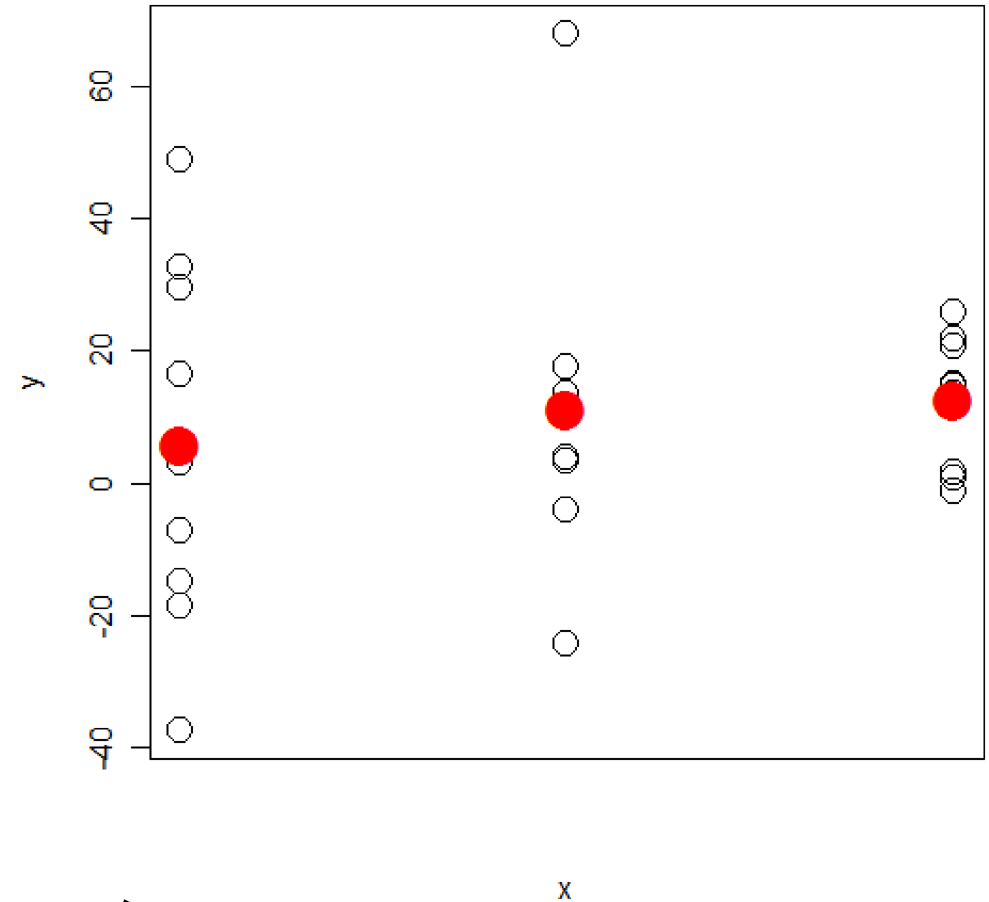
**⇒ なんでこんなことをするの？**

# 分散分析の考え方　２

## 絶対差ある

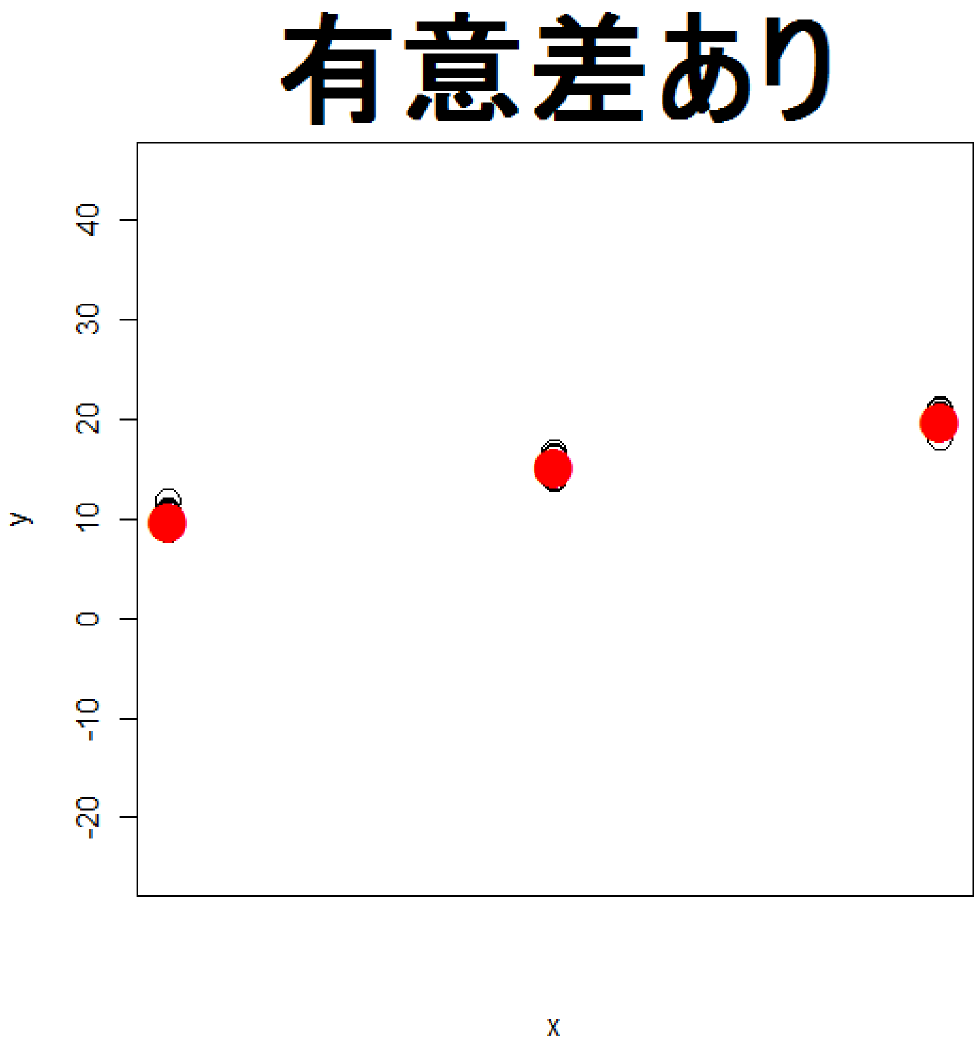


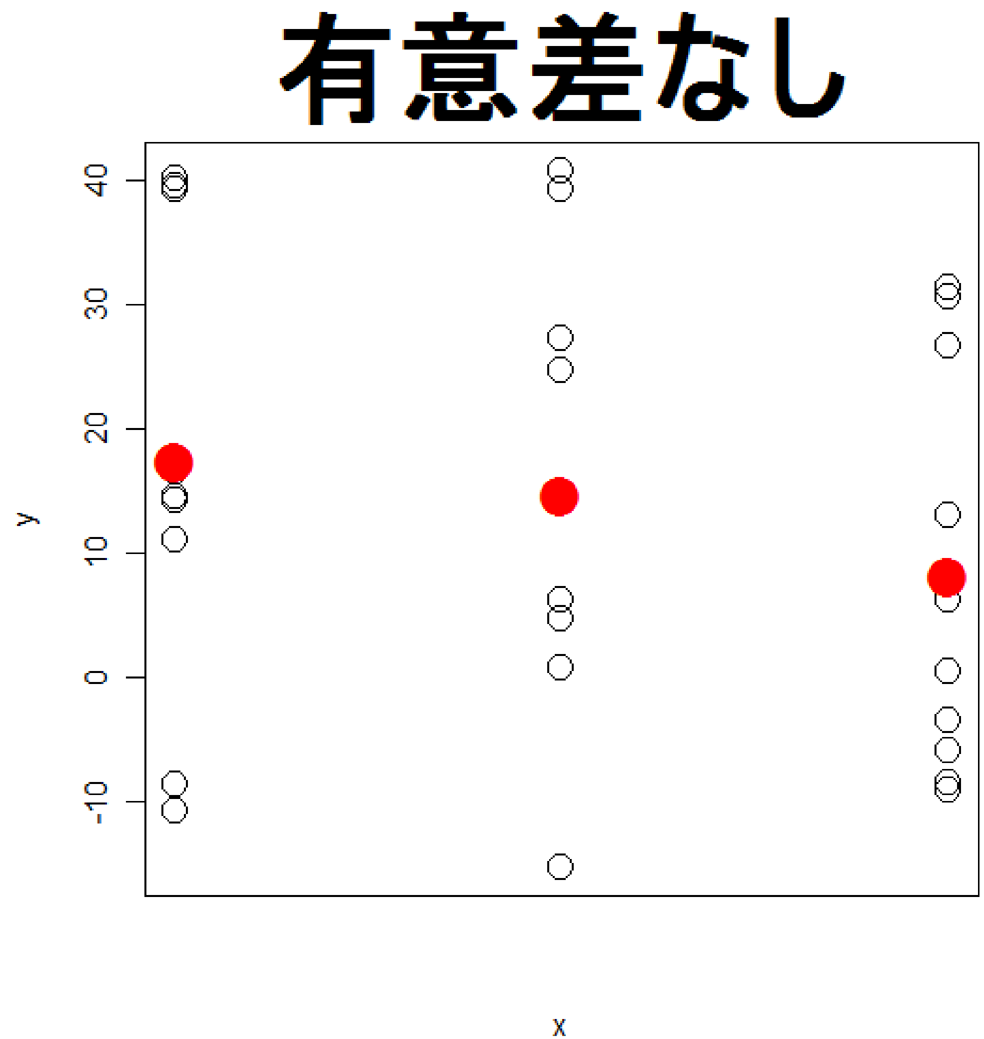

## 絶対差ない




この違いは何か？

# 平均値の差はほとんど一緒だけど…



# なぜ、平均値の差が小さくても有意？

## Ans. <u>誤差が小さいから</u>

- 0.001mm 単位で測れる電子顕微鏡

⇒1mm の差でも有意になる

- ボロボロな定規

⇒1cm 程度の差でも有意じゃない

⇒ 差が大きいからと言って有意とは限らない

# いつ「意味が有る差」になるの？

差が大きくても有意じゃないかも

差が小さくても有意かも

有意・有意じゃないの基準は……？

$$\frac{\text{効果}}{\text{誤差}}$$

# 誤差と効果

$$\frac{\text{効果}}{\text{誤差}}$$

- この値（Ｆ比）が大きい

⇒ 誤差に比べて効果が大きい

有意！

- この値が小さい

⇒ 効果より誤差がでかい

有意じゃない（差は誤差の範囲内）

# 分散分析の目的

データを「誤差」と「効果」に分ける

で、

$$\frac{\text{効果}}{\text{誤差}}$$

を計算する

この値を見て有意か有意じゃないかを判断

# 分散と言う " 威力 "

「効果」とか「誤差」はどうやって定量化？

## <u>分散</u>

# なぜ分散か

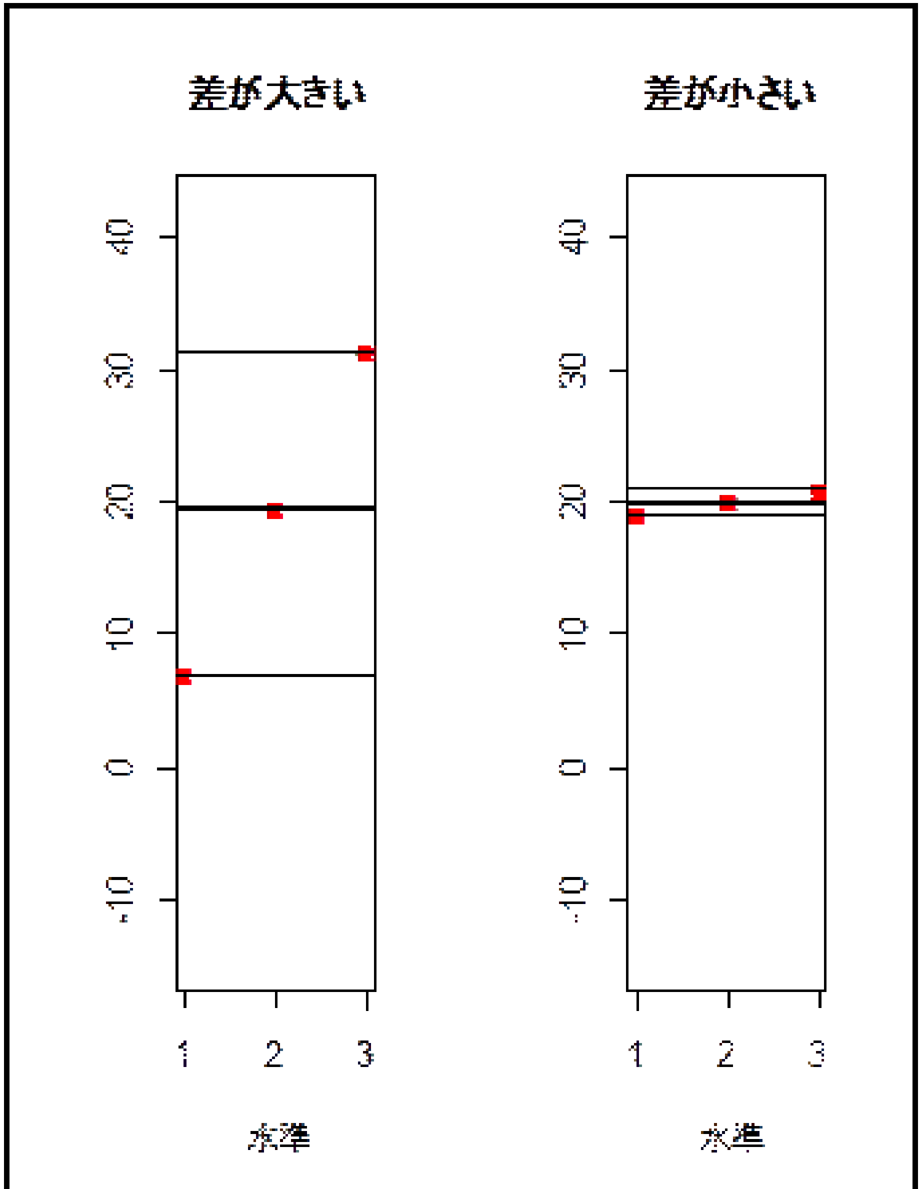
分散分析の目標

⇒ 平均値の差を見る

- 分散は……

<u>平均値の差の大きさを表してくれる</u>

# 分散という"威力"




平均値の差が大きい

↓

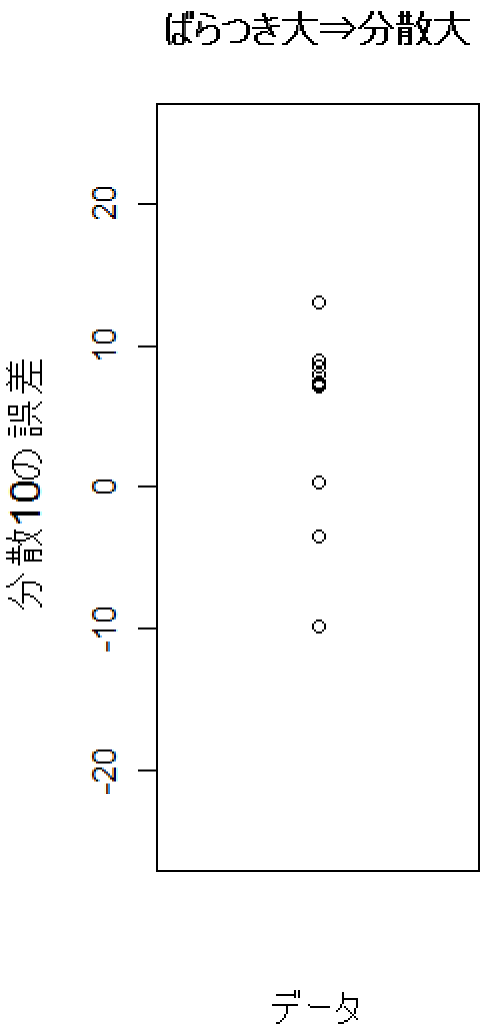
平均値同士の「ばらつき」が大きい

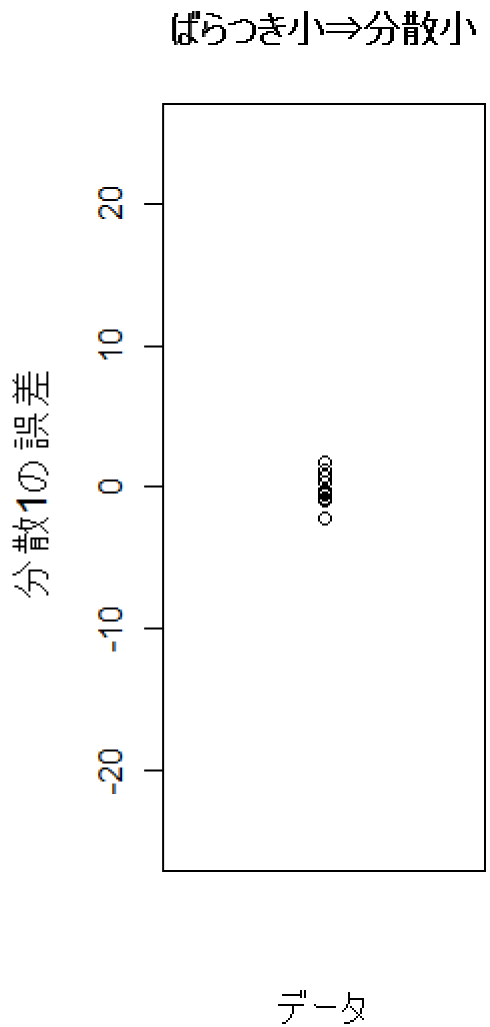
⇒ 分散が大きい

↓

分散大で、平均値の差は大

# 誤差の威力




誤差が大きい

↓

意味不明な「ばらつき」が大きい

↓

誤差の分散が大きい

# まとめ

- 分散分析は

$$\frac{\text{効果}}{\text{誤差}}$$

- を計算して、有意かどうかを判断する
- 誤差とか、効果は「分散」で表すことができる
- だから、分散分析では、「平均値同士の差の分散」と「誤差の分散」の比を取って統計量とする
- p 値は、F 分布を使うと計算可能

# 分散分析の計算

# 分散分析の計算

データを

## 平均値 ＋ 効果 ＋ 誤差

に分ける

で、効果と誤差を分離する

# 問い

魚の体長は薬 A,B,C によって「有意に」変わるか？

- 帰無仮説：違いは誤差の範囲内
- 対立仮説：誤差じゃない
- p 値： A,B,C により違いがあるように見えるんだけれども、実はそれは単なる誤差による違いなのだよと言える確率
- 有意水準　5%

| | A | B | C |
|---|---|---|---|
| 1 | 13 | 11 | 8 |
| 2 | 11 | 10 | 7 |
| 3 | 6 | 7 | 5 |
| 4 | | 7 | 5 |
| 5 | | 5 | 4 |
| 6 | | | 3 |
| 7 | | | 3 |
| 列の合計 | 30 | 40 | 35 |
| 列の平均 | 10 | 8 | 5 |
| 列の効果 | 3 | 1 | -2 |

# 分散分析の計算

総平均

7

# 効果

| | A | B | C |
|---|---|---|---|
| 1 | 3 | 1 | -2 |
| 2 | 3 | 1 | -2 |
| 3 | 3 | 1 | -2 |
| 4 | | 1 | -2 |
| 5 | | 1 | -2 |
| 6 | | | -2 |
| 7 | | | -2 |

# 効果

- 効果に全平均(7)を足すと、列の平均（ 10,8,5 ）になる
- Aの効果は3

  ⇒Aは”並”な魚(体長7 )に3だけ上方修正させる効果がある
- Cの効果は－ 2

  ⇒” 並”な魚を－ 2だけ下方修正させる効果がある

| | A | B | C |
|---|---|---|---|
| 1 | 3 | 1 | -2 |
| 2 | 3 | 1 | -2 |
| 3 | 3 | 1 | -2 |
| 4 | | 1 | -2 |
| 5 | | 1 | -2 |
| 6 | | | -2 |
| 7 | | | -2 |

# 誤差

| | A | B | C |
|---|---|---|---|
| 1 | 3 | 3 | 3 |
| 2 | 1 | 2 | 2 |
| 3 | -4 | -1 | 0 |
| 4 | | -1 | 0 |
| 5 | | -3 | -1 |
| 6 | | | -2 |
| 7 | | | -2 |

# 誤差

”並”な魚に薬の効果を足したら、魚の体長が出るはず

⇒ でも、実際は、理論値からぶれている

**<u>誤差</u>**

データ = 平均 + 効果 + 誤差
　より

誤差 = データ - 平均 - 効果
　で計算可能

| | A | B | C |
|---|---|---|---|
| 1 | 3 | 3 | 3 |
| 2 | 1 | 2 | 2 |
| 3 | -4 | -1 | 0 |
| 4 | | -1 | 0 |
| 5 | | -3 | -1 |
| 6 | | | -2 |
| 7 | | | -2 |

# 分散の計算方法

”並”な魚の平均体長(７)

薬による上方(下方)修正

誤差によるばらつき　　　　　　　　に分けられた

- 効果による平方和

誤差による平方和

⇒ さっきの表を２乗して足すだけ

分散＝　平方和÷自由度

# 効果の自由度

合計０

↓

１個固定

↓

A,B,C のうち、
１個固定

３ - １ = 2

| | A | B | C |
|---|---|---|---|
| 1 | 3 | 1 | -2 |
| 2 | 3 | 1 | -2 |
| 3 | 3 | 1 | -2 |
| 4 | | 1 | -2 |
| 5 | | 1 | -2 |
| 6 | | | -2 |
| 7 | | | -2 |

# 誤差の自由度

各列ごとに和が０

↓

３つ固定

↓

１５－３＝１２

| | A | B | C |
|---|---|---|---|
| 1 | 3 | 3 | 3 |
| 2 | 1 | 2 | 2 |
| 3 | -4 | -1 | 0 |
| 4 | | -1 | 0 |
| 5 | | -3 | -1 |
| 6 | | | -2 |
| 7 | | | -2 |

# 結果

効果の平方和　６０

効果の自由度　２

★効果の分散　６０÷２＝３０

誤差の平方和　７２

誤差の自由度　１２

★誤差の分散　７２÷１２＝６

Ｆ比

＝効果÷誤差

＝３０÷６

＝５

ｐ＝０．０２６

有意差あり

# R の結果　1

```
> aov.model<-aov(Tl~Factor,data=data)
> summary(aov.model)
             Df   SumSq   MeanSq  Fvalue  Pr(>F)
Factor        2   60        30       5    0.02634 *
Residuals    12   72         6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> summary(lm.model)

Call:

lm(formula = TI ~ Factor, data = data)

Residuals:

  Min     1Q Median     3Q    Max

 -4.0   -1.5    0.0    2.0    3.0

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)   10.000      1.414   7.071  1.3e-05 ***

FactorB       -2.000      1.789  -1.118    0.285

FactorC       -5.000      1.690  -2.958    0.012 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.449 on 12 degrees of freedom

Multiple R-squared: 0.4545,     Adjusted R-squared: 0.3636

F-statistic:     5 on 2 and 12 DF,  p-value: 0.02634
```

# 教科書の内容

# 教科書内容　1

- モデル式

WEIGHT=SEX

・・・ Rでは「～」を使う

WEIGHT~SEX・ TI ～ Factor・ 体長～薬

- 一般線形モデル

回帰分析とか分散分析とかは皆まとめて一般線形モデル（ General Linear Model ）

# 教科書内容　２

- 変動の分解

SSY　全平方和　　　　　（生データのばらつき）

SSE　誤差平方和　　　（誤差のばらつき）

SSF　薬による平方和　（効果のばらつき）

- SSY=SSE+SSF

データ＝並な魚の体長＋薬＋誤差

⇒ 並な魚は個性なし⇒ばらつきなし（すべて７）

# 教科書内容　３

- 分散分析の仮定
  ⇒ 等分散正規分布

- 満たしてないと使えない
  （Ｆ検定は有る程度頑健らしい？）

# 実演 with R