実験データー解析概論

— 統計学に基づく「よりよい推論」のために —

三中 信宏

MINAKA Nobuhiro

独立行政法人 農業環境技術研究所 生態系計測研究領域 上席研究員［生物統計学］

東京大学大学院 農学生命科学研究科 生物・環境工学専攻 教授［生態系計測学］

東京農業大学大学院 農学研究科 客員教授［応用昆虫学］

mailto:minaka@affrc.go.jp （メール）

http://twitter.com/leeswijzer/ （ツイッター）

http://cse.niaes.affrc.go.jp/minaka/ （ウェブサイト）

http://d.hatena.ne.jp/leeswijzer/ （ブログ）

2014 年度・東京農業大学生物応用化学科（毎週金曜第 2 限）

三中信宏〈実験データー解析概論〉：2014 年 4 月～ 2014 年 9 月

東京農業大学世田谷キャンパス 431 番教室

実験計画法

実験区の配置と分散分析に関する理論

実験要因の処理効果の有無を判定し，処理区間のちがいを統計学的に判定する．

１）【反復実施】：ある処理のばらつきを評価する．

２）【無作為化】：背景要因の影響を誤差に転化する．

３）【局所管理】：実験区全体を分割しブロック化する．

実験計画法

実験区の割付けにはじまり統計分析におわる

［1］実験区の割付け（レイアウト）

［2］統計分析（線形統計モデル）

2-1）分散分析

2-1-1：正規性・等分散性の確認

2-1-2：一般化線形モデル，ノンパラメトリック法

2-2）多重比較

処理平均の比較と補正，信頼区間の構築

1 要因完全無作為化法（Box 1）

［実験］殺虫剤散布処理（7 水準 a ～ g）によるイネ収量実験を 4 反復の完全無作為化法で実施した（IRRI）.

a	b	c	d	g	f	c
e	d	a	g	c	g	f
d	f	g	b	e	d	b
b	c	e	a	f	a	e

「完全無作為化法」とは，圃場全体にわたって実験区を無作為化する割付けである.

1 要因完全無作為化法（Box 1）

線形モデルと分散分析への道

データ　総平均　誤差項

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (\epsilon_{ij} \sim N(0, \sigma^2))$$

第 i 水準
第 j 標本

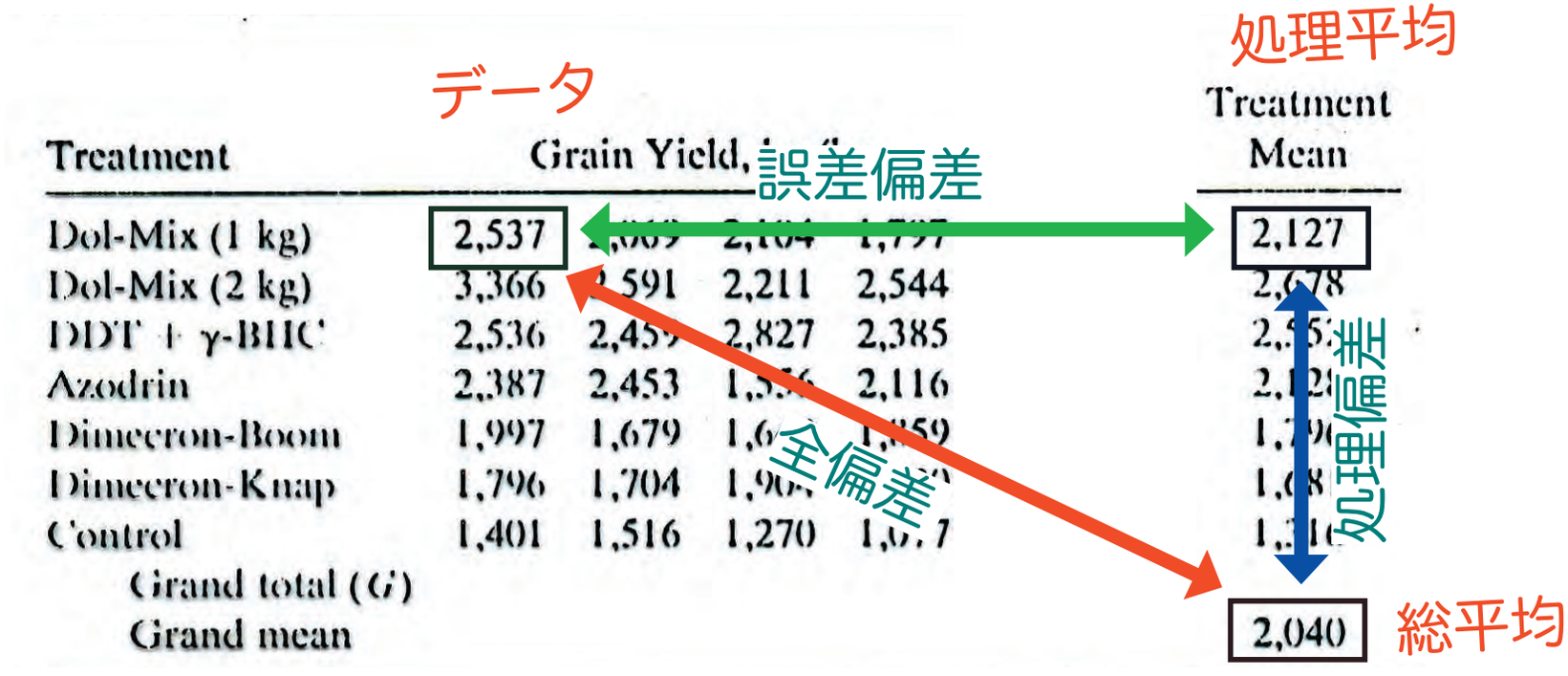
処理効果

誤差の正規性（仮定）

データに影響する変動因は「処理効果」と「誤差効果」の二つだけだから，データのばらつきを処理効果に由来する部分と誤差効果に由来する部分に分割する．

1 要因完全無作為化法（Box 1）

偏差を分割する（1）

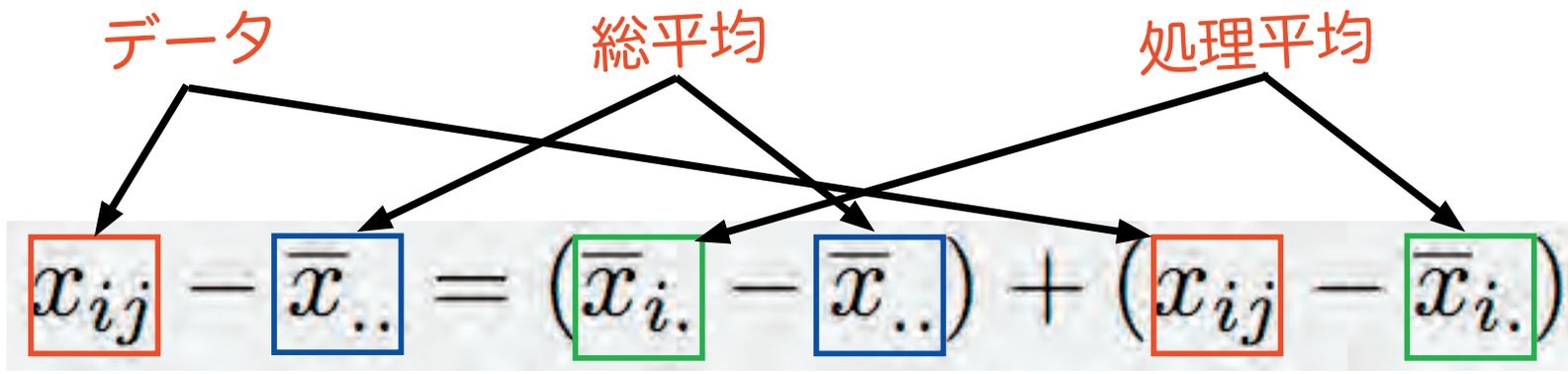


Treatment	Grain Yield,			Treatment Mean
Dol-Mix (1 kg)	2,537	[illegible]	[illegible]	2,127
Dol-Mix (2 kg)	3,366	[illegible]591	2,211	2,544	2,678
DDT + γ-BHC	2,536	2,459	2,827	2,385	2,5[illegible]
Azodrin	2,387	2,453	1,5[illegible]	2,116	2,1[illegible]
Dimecron-Boom	1,997	1,679	1,6[illegible]	[illegible]59	1,7[illegible]
Dimecron-Knap	1,796	1,704	1,90[illegible]	[illegible]	1,6[illegible]
Control	1,401	1,516	1,270	1,[illegible]7	1,[illegible]
Grand total (G)					
Grand mean					2,040

「全偏差＝処理偏差＋誤差偏差」と分割する

1 要因完全無作為化法（Box 1）

偏差を分割する（2）



総平均に対する各データの偏差
（総効果）

総平均に対する各処理平均の偏差
（処理効果）

各処理平均に対するデータの偏差
（誤差効果）

1 要因完全無作為化法（Box 1）

平方和を分割する（1）

全平方和

$$\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = \sum_i \sum_j \{(\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.})\}^2$$

処理偏差　　誤差偏差

$$= \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$$

処理平方和　　誤差平方和

$$+ 2 \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})(x_{ij} - \bar{x}_{i.})$$

ここで：

$$\sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})(x_{ij} - \bar{x}_{i.}) = \sum_i (\bar{x}_{i.} - \bar{x}_{..}) \sum_j (x_{ij} - \bar{x}_{i.})$$

$$= 0$$

1 要因完全無作為化法（Box 1）

平方和を分割する（2）

したがって：

$$\sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = \sum_i \sum_j (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$$

全平方和　　処理平方和　　誤差平方和

「全平方和＝処理平方和＋誤差平方和」と分割できた

1 要因完全無作為化法（Box 1）

自由度を分割する

1)「全自由度」は，データの偏差に対する総平均の制約がひとつあるので，「$7 \times 4 - 1 = 27$」；

2)「処理自由度」は，処理平均の偏差に対する総平均の制約がひとつあるので，「$7 - 1 = 6$」；

3)「誤差自由度」は，データの偏差に対する 7 つの処理平均の制約があるので，「$7 \times 4 - 7 = 21$」.

「全自由度＝処理自由度＋誤差自由度」と分割できた

1 要因完全無作為化法（Box 1）

平均平方（分散）を計算する

水準数が t，反復数が r であるとき，処理平均平方は：

$$\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}(\overline{x}_{i.}-\overline{x}_{..})^2}{t-1}=\frac{r\sum_{i=1}^{t}(\overline{x}_{i.}-\overline{x}_{..})^2}{t-1}$$

誤差平均平方は：

$$\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}(x_{ij}-\overline{x}_{i.})^2}{tr-t}$$

1 要因完全無作為化法（Box 1）

分散比 F 値と F 検定（1）

処理平均平方と誤差平均平方との比（F 値）は，誤差効果に対する処理効果の相対的な大きさを表す.

$$F = \frac{r\sum_{i=1}^{t}(\overline{x}_{i.} - \overline{x}_{..})^2/(t-1)}{\sum_{i=1}^{t}\sum_{j=1}^{r}(x_{ij} - \overline{x}_{i.})^2/(tr-t)}$$

直感的には，F 値が大きいほど処理水準間には「差がある」と認知されるが，その明確な基準は F 検定が与える.

1 要因完全無作為化法（Box 1）

分散比 F 値と F 検定（4）

帰無仮説 H_0「処理効果はない」のもとで，関係する統計量は次のような確率分布をする：

$$x_{ij} = \mu + \epsilon_{ij} \sim N(\mu, \sigma^2)$$

$$\frac{r\sum_{i=1}^{t}(\overline{x}_{i.} - \overline{x}_{..})^2}{\sigma^2} \sim \chi^2(t-1)$$

$$\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}(x_{ij} - \overline{x}_{i.})^2}{\sigma^2} \sim \chi^2(tr-t)$$

処理効果と誤差効果に関する平方和は互いに独立にカイ二乗分布する.

1 要因完全無作為化法（Box 1）

分散比 F 値と F 検定（5）

互いに独立なカイ二乗変量をその自由度で割ったものの比は F 分布をする．したがって，帰無仮説 H_0 のもとで，下記の F 値は F 分布に従う．

$$\frac{\frac{r\sum_{i=1}^{t}(\bar{x}_{i.}-\bar{x}_{..})^2}{\sigma^2}/(t-1)}{\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i.})^2}{\sigma^2}/(tr-t)} \sim F(t-1,tr-t)$$

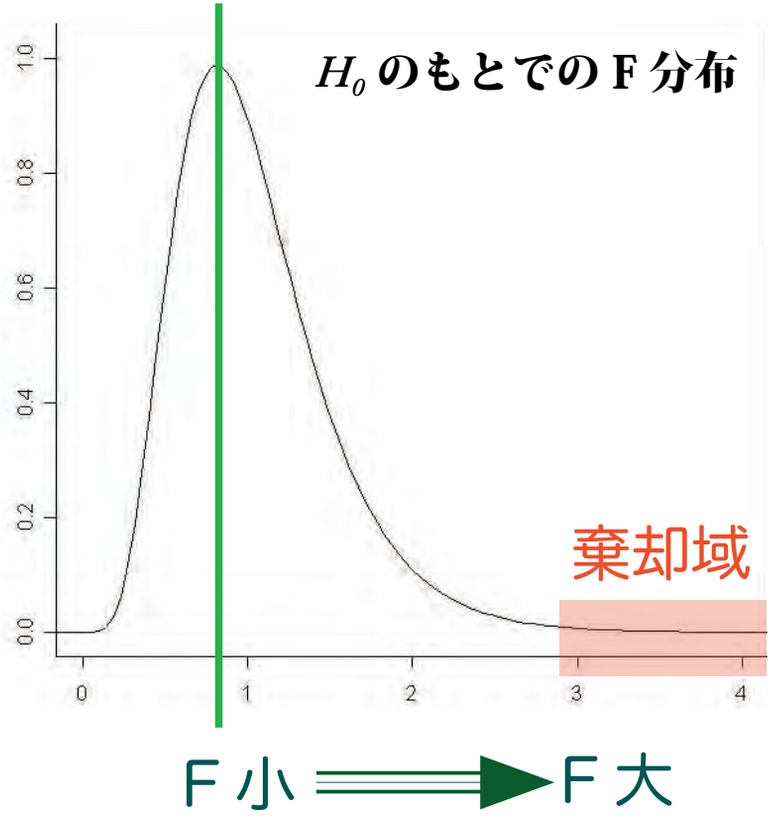
$$\frac{r\sum_{i=1}^{t}(\bar{x}_{i.}-\bar{x}_{..})^2/(t-1)}{\sum_{i=1}^{t}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i.})^2/(tr-t)} \sim F(t-1,tr-t)$$

この F 値は，「処理平均平方／誤差平均平方」の比であることに注意．

1 要因完全無作為化法（Box 1）

分散比 F 値と F 検定（6）

帰無仮説 H_0 のもとで F 分布のグラフにおける棄却域を設定する.



帰無仮説 H_0 のもとでの F 値の期待値は 1 に近い．データから得られた F 値が，棄却域（上側 5% あるいは 1%）に入るほど大きな値であるとき，「効果がない」とする帰無仮説 H_0 を棄却して，「効果がある」とする対立仮説 H_1 を受容する.

1 要因完全無作為化法（Box 1）

分散分析表

変動因	自由度	平方和	平均平方	F 値	F(5%)	F(1%)
全体	27	7,577,412				
処理	6	5,587,175	931,196	9.8255**	2.57	3.81
誤差	21	1,990,237	94,773		F(6,21)	

上の結果から，Box 1 の殺虫剤散布実験では，処理要因の効果は 1% レベルで有意であることが判明した.