

統計對農業研究的貢獻

呂椿棠

農業委員會農業試驗所

101.9.5

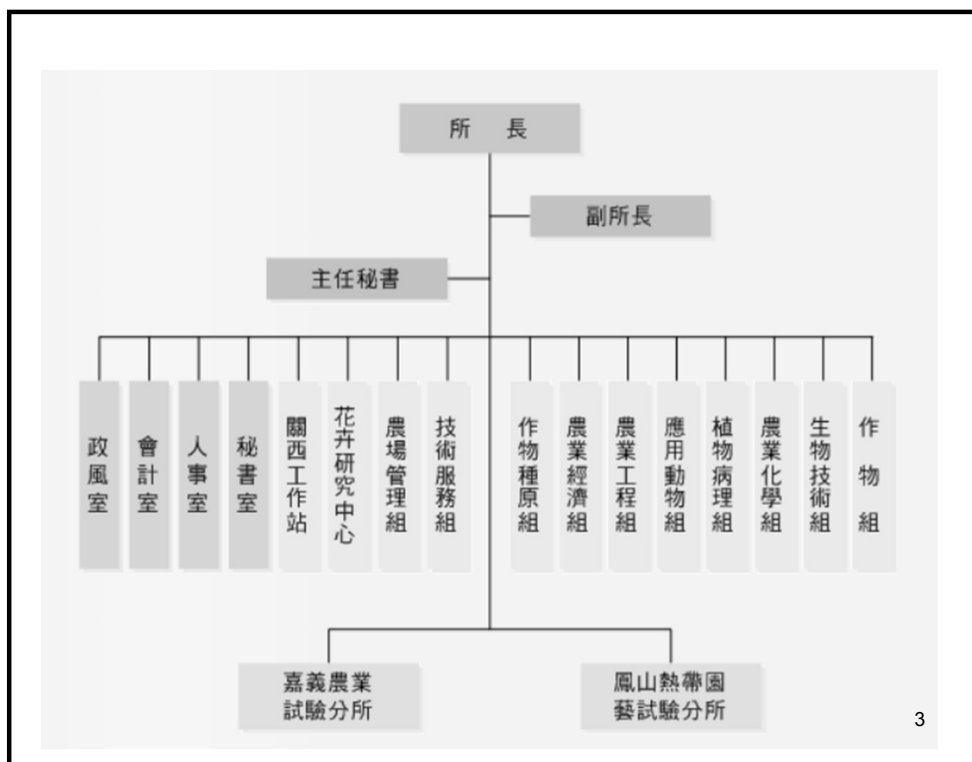
行政院農業委員會農業試驗所簡介

本所創立於民國前十七年，原稱「台灣總督府農業試驗所」，民國三十四年台灣光復，改稱「台灣省農業試驗所」，直隸行政長官公署。三十八年改隸台灣省政府農林廳，八十八年七月一日隸屬行政院農業委員會。

本所原設於台北市羅斯福路四段，由於都市發展，環境污染，且面積狹小，不足因應現代農業試驗研究之需，因此於民國六十六年十二月遷移台中縣霧峰鄉萬豐村現址。

本所佔地145公頃，其中建築面積17公頃，主要建築包括行政大樓，試驗研究大樓及各種附屬建築等。農場面積128公頃，灌溉排水設施完善，為全國首善之農業試驗研究環境。





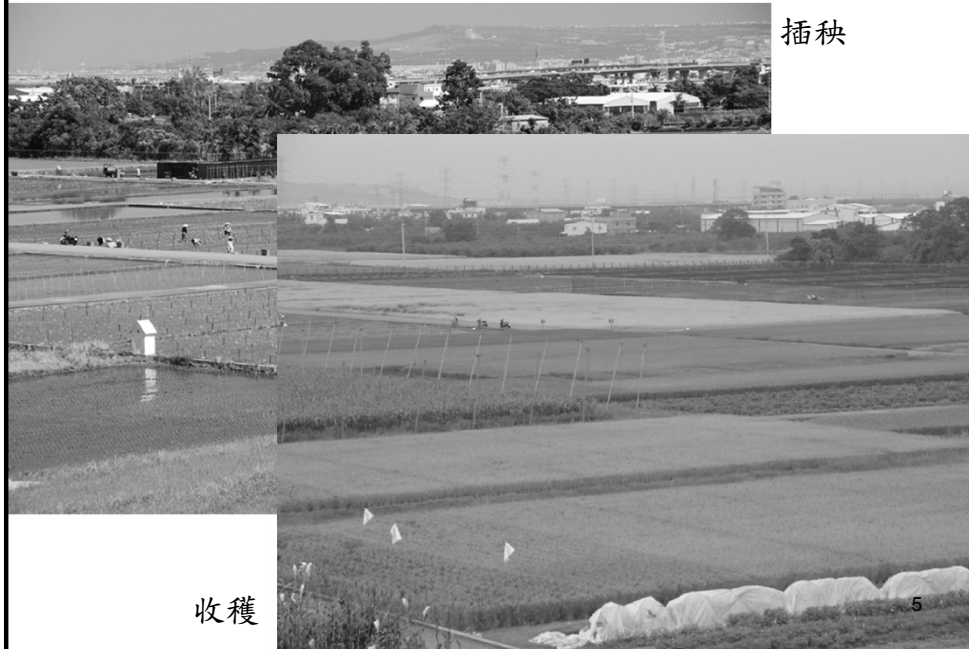
3

研究簡介

- ☞ 本所從事農藝與園藝作物之遺傳、育種、種原、生理、組織培養、生物科技、土壤肥料、植物營養、病蟲害防治、農業工程及氣象、農業經濟、農業推廣及農場經營管理等之科技研究，以改進各種作物之產量與品質。
- ☞ 作物組生物統計與生物資訊研究室，成立於1977年原為農藝系試驗統計研究室，為國內農業研究機關第一個且目前唯一的生物統計與生物資訊學術研究單位。其宗旨在有效發揮農業試驗設計與資料分析的合作與諮詢功能，促使各項推動之研究計畫達成目標，以提昇國內農業研究的效率及水準。

4

農業試驗所試驗田區一隅



插秧

收穫

5

生物統計學的演變

演化論 → 優生學 → 敘述統計學 → 生物統計學

Darwin
1895年
種的起源

Galton
1870年
遺傳的天才

K.Pearson
相關,迴歸

Pearson分佈曲線
偏相關,重相關
動差,卡方檢定

細胞學 → 遺傳學 → 田間實驗 → 實驗設計學 → 推測統計學

1865年
發表Mendel定理
1900年
再發現Mendel定理

R.A.Fisher
1920年
Rothawsted
試驗場

R.A.Fisher
重複
隨機化
局部控制
RCBD
Latin Method
ANOVA

擬定假設 → 實驗 → 假說檢定
Gosset (student) t分布
Fisher 實驗設計法
E.S. Pearson
Neyman } 擬說檢定

6

影響農業研究的統計大師-Fisher

- ☞ Sir Ronald Aylmer Fisher, 1890~1962 ; Fisher 1912年通過了數學學位考試，並且在隔年從劍橋大學畢業。
- ☞ 他是二十世紀初期的一位著名統計學家，對於統計學的貢獻許多，奠定了近代統計學蓬勃發展的基石，學過統計學的人不可能沒聽過他的名字。他更是現代試驗設計的鼻祖，對於驗證一個科學假說時，要如何設計實驗、採集數據，才能避開不必要的干擾因素、可靠地回答我們想問的問題，他即是這一切方法的開山祖師。

7

- ☞ 西元1919年春天，Fisher 應聘到倫敦北部的Rothamsted Experimental Station從事試驗數據的分析工作。除了在其中擔任一名統計員之外，所長John Russell（約翰·羅素）也讓他設立了一個統計實驗室。
- ☞ Fisher曾說：「最初幾年，我唯一能做的事，就是設法讓那些研究人員了解為什麼無法從他們的試驗數據分析出有價值的結論」。在1920及1930年代期間，Fisher致力於試驗設計與統計分析的研究並發表了一系列不朽的著作。其中，尤其是在1925年出版的Statistical Methods for Research Workers(研究者的統計方法)與在1935年出版的The Design of Experiments(實驗設計)兩部巨著，以及在1926年發表的論文The Arrangement of Field Experiments(田間試驗佈置)對設驗設計學的發展影響最為深遠。

8

- ☞ Fisher此段期間，也為歐美各地、各方面的研究人員，直接或間接地提供了許多試驗統計上的諮詢服務。由於他的卓越貢獻，終於使得各方面的研究人員體認到；除非先取得有價值的試驗資料，否則就無法在資料分析中獲致有價值的結論；而為取得有價值的資料，除非遵循統計原理先設計出有效率的試驗計畫，並依計畫實施不可。也因此使得一向只扮演著「試驗工作之善後處理者」的統計人員，在試驗性研究工作的領域裡建立起「由始至終的積極參與者」之地位。
- ☞ 在試驗統計學發展之初，Fisher, Yates, Snedecor及Cochran等學者大多把研究的重心投注於農學研究的試驗設計與統計分析方法上；而他們所創立的基本學理與方法卻迅速而廣泛地被延用至其他各門各類的試驗性研究工作中。近年來，由於科學研究領域的不斷擴展，衍生出許多試驗統計上的新問題，因而使許多學者轉而致力於醫學、教育科學、環境科學及工商科學等之試驗統計方法的開發。

9

統計學的重要性

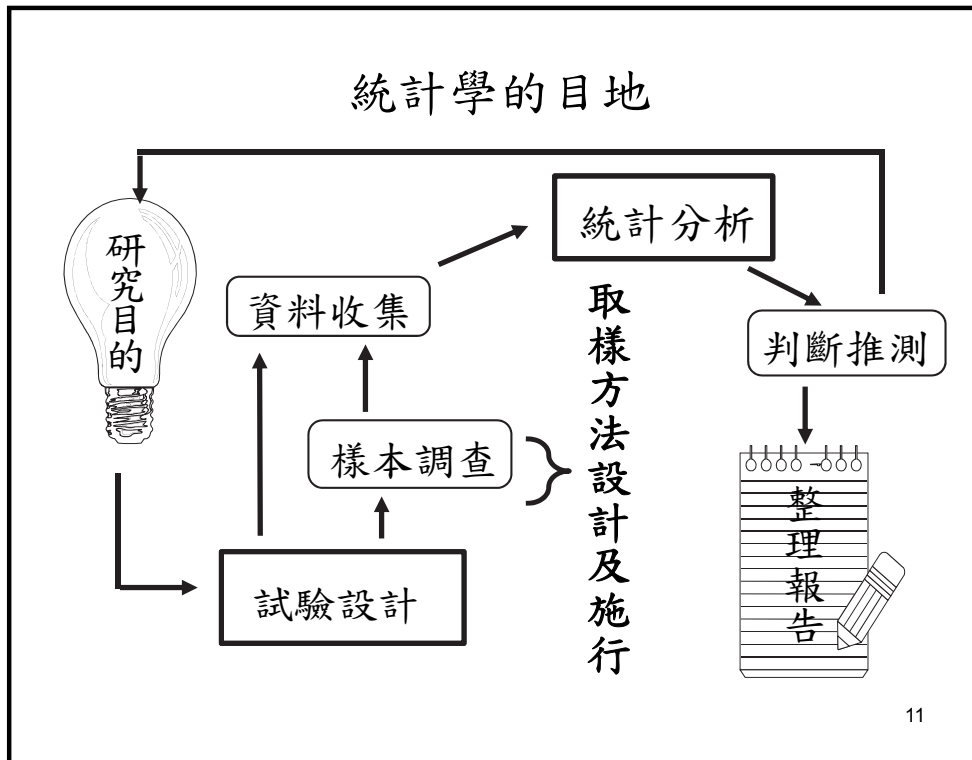
What? 統計學之目的：從數據中找訊息

HOW?

1. 如何產生數據 ← 試驗設計、取樣
2. 如何整合數據 ← 資料分析
3. 怎樣從數據中得到結論 ← 統計推論

所謂統計學，是在包含不確定性的現實中，研究如何利用訊息來思考並給出行動方針的學問

10



統計學主要可分成：

敘述統計(descriptive statistics)

將資料組織整理，加以綜合描述，只蒐集一族群（population，母體）中的資料並以該族群的資料對該族群的特性下結論

推測統計(inferential statistics)

利用各種測度解釋樣本(sample，樣品)資料以推測族群特性

⇒ { 介量(parametric)
非介量(non-parametric)

12

試驗性研究之特質-演繹與歸納

- ☞ 在科學活動中，研究工作者推理的方式大致可分為演繹式推理(deductive inference)與歸納式推理(inductive inference)兩種。所謂演繹式推理就是根據
- ☞ (1)對於研究對象之整體(統計學稱之為族群或母體(population))已有的知識，
- ☞ 或(2)已經獲得確證的定律或不證自明的公理等，
- ☞ 而對研究對象之特定的局部個體所做的推論。這種推理方式概用之於以既經驗證或已被接受的學說為基礎，推衍出一些必然的結果，從而拓展成有待驗證的新假說。而所謂歸納式推理則是根據在研究對象之局部個體(統計學稱之為樣品(sample))上的發現，而對研究對象之整體所做的推論。這種推理方式概用於新知的發掘與收集，從而創造理論來說明所發現的通則；或是用之於追索客觀的驗證以支持新的假說或修正既存的學說。

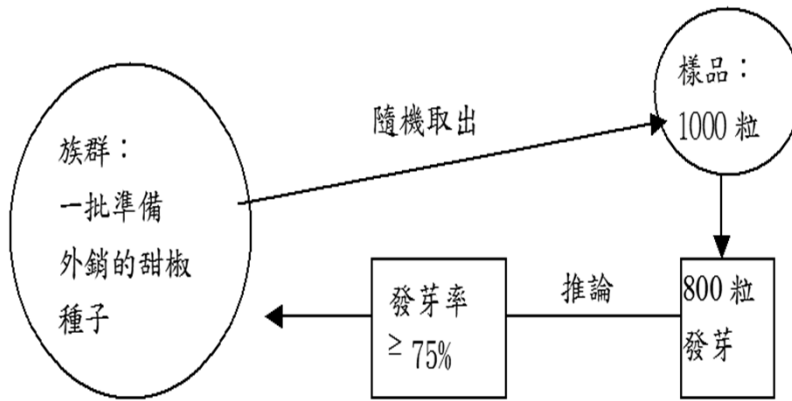
13

例釋如下：

- ☞ 假設我國農業主管機關明訂外銷用蔬菜種子之發芽率的最低標準為75%；而其誤差之容忍範圍為5%。現在，假設有一批甜椒種子準備外銷。檢驗人員決定從中隨機取出1000粒種子進行發芽試驗；並決定：若被抽出的1000粒種子中發芽的粒數小於或等於726，就判定該批種子的發芽率不合格；而若發芽粒數大於726就判定該批種子的發芽率合乎外銷標準。在完成發芽試驗後，檢驗人員發現被抽出的1000粒種子中有821粒發芽良好，據此檢驗人員判定該批種子之發芽率合乎標準。
- ☞ 在本例，由於隨機抽出的1000粒種子中會發芽的粒數(X)是一個呈二項分布的隨機變數；據此，檢驗人員可以在這批種子的發芽率為75%以上或70%以下的兩種假設上，分別演繹出下列兩個必然的結論：

14

樣品取樣與估計示意圖



15

- 如果發芽率確在75%以上，則在隨機取出的1000粒中，發芽粒數小於或等於726的機率低於5%。

$$\Pr[X \leq 726 | P \geq 0.75] \leq \sum_{x=0}^{726} \binom{1000}{x} (0.75)^x (0.25)^{1000-x} \approx 0.0431$$

- 如果發芽率確在70%以下，則在隨機取出的1000粒中，發芽粒數大於726的機率低於4%。

$$\Pr[X > 726 | P \leq 0.70] \leq \sum_{x=727}^{1000} \binom{1000}{x} (0.70)^x (0.30)^{1000-x} \approx 0.0337$$

- 其次，檢驗人員根據被抽出的1000粒甜椒種子中有821 (>726) 粒發芽成功的事實，歸納出這批甜椒種子之發芽率在75%以上(合乎外銷標準)的結論。

16

- ☞ 這種歸納性結論的正確性只是「或然」而非「必然」的。因為：縱使這批種子之發芽率確在70%以下，仍有可能得到與上述完全相同的試驗結果，而使它僥倖地被判定為合乎標準的種子。類此，試驗性研究過程中，研究者固然會交替地運用演繹推理與歸納推理；但在本質上，試驗性研究均為一種抽樣調查的過程，而後根據在樣品上收集到的資訊對族群做成歸納性的結論。因而試驗性研究所獲得的結論並不必然就是真理。
- ☞ 試驗設計之最主要的功能就是依據統計學原理，在試驗之前擬定正確的抽樣的方法與最經濟的抽樣量，以使試驗後所獲致的歸納性結論之正確性具有在既定水準以上的機率保證。

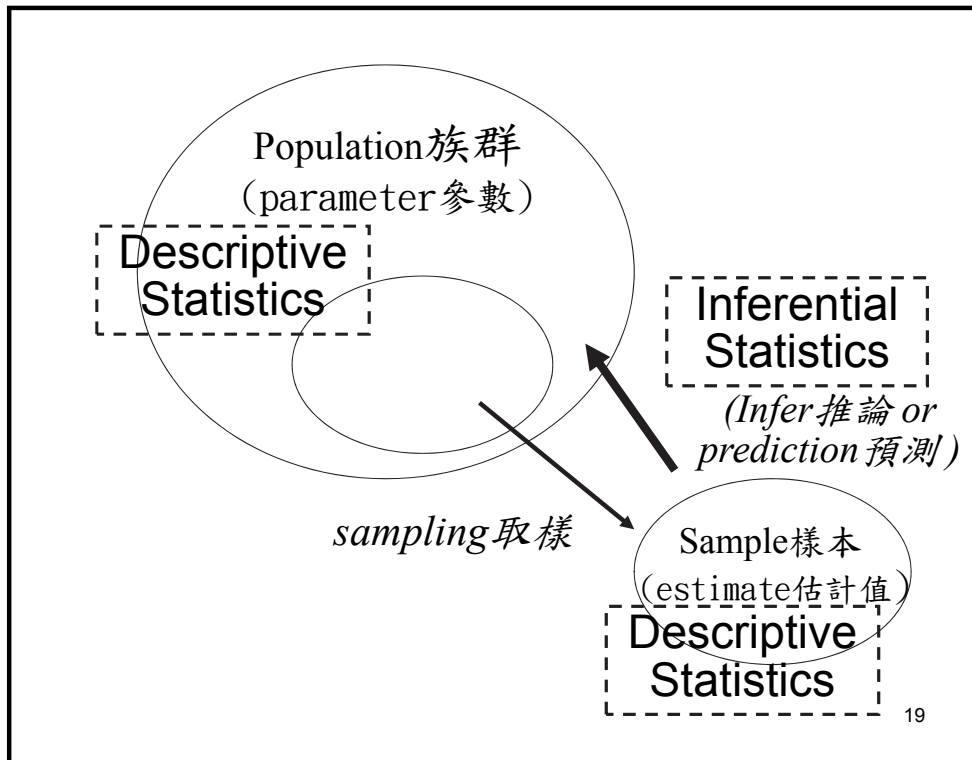
17

結論錯誤的類型

- ☞ 在上面釋例這種「品管式」的試驗裡，其試驗之結論有可能發生下述兩者之一的錯誤：
- ☞ (1)將發芽率合乎標準的種子誤判為不合乎標準而禁止外銷，因而造成種子生產者投資上的損失。
- ☞ (2)將發芽率低於70%以下的種子誤判為合乎標準而核准外銷，因而造成購買者要求賠償或影響商譽等的損失。
- ☞ 前者就是統計學之假說檢驗(hypothesis testing)中所謂的第I型錯誤(type I error)，發生這種錯誤的機率習慣上以 α 表示之；而後者即所謂之第II型錯誤(type II error)，習慣上以 β 表示其發生之機率，示如下表：

試驗後之結論。	潛在的真相。	
	發芽率 $\geq 75\%$ 。	發芽率 $< 75\%$ 。
發芽率 $\geq 75\%$ 。	結論正確。	第II型錯誤。 (發生機率 β)。
發芽率 $< 75\%$ 。	第I型錯誤。 (發生機率 α)。	結論正確。

18



檢定與推估

檢定(Testing of hypothesis)

虛無擬說 $\xrightarrow{\text{判斷}}$ 拒絕或不拒絕

欲解決問題：「介量是否等於某個特定值？」

推估(Estimation)

樣本 $\xrightarrow{\text{推測}}$ 族群之平均和變異大小

欲解決問題：「族群介量的值是什麼？」

統計上的假說(hypothesis)

例：水稻育種者擬比較4個新品種水稻之間，其稻穀產量是否有差異？

虛無擬說(null hypothesis)

（因為此假說有被推翻之可能，故R.A.Fisher稱之為虛無擬說，即無效假說）

H_0 ：4個品種的稻穀產量相同

$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = \bar{x}_4$$

對立擬說(alternative hypothesis)

H_1 ：4個品種的稻穀產量至少有兩個不等

21

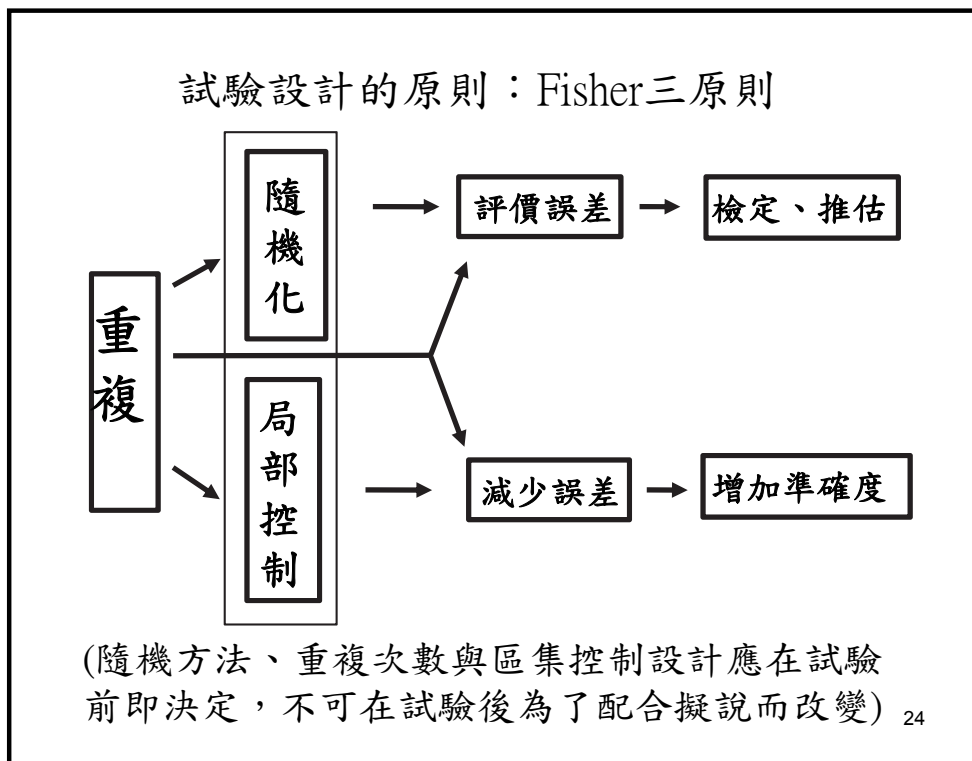
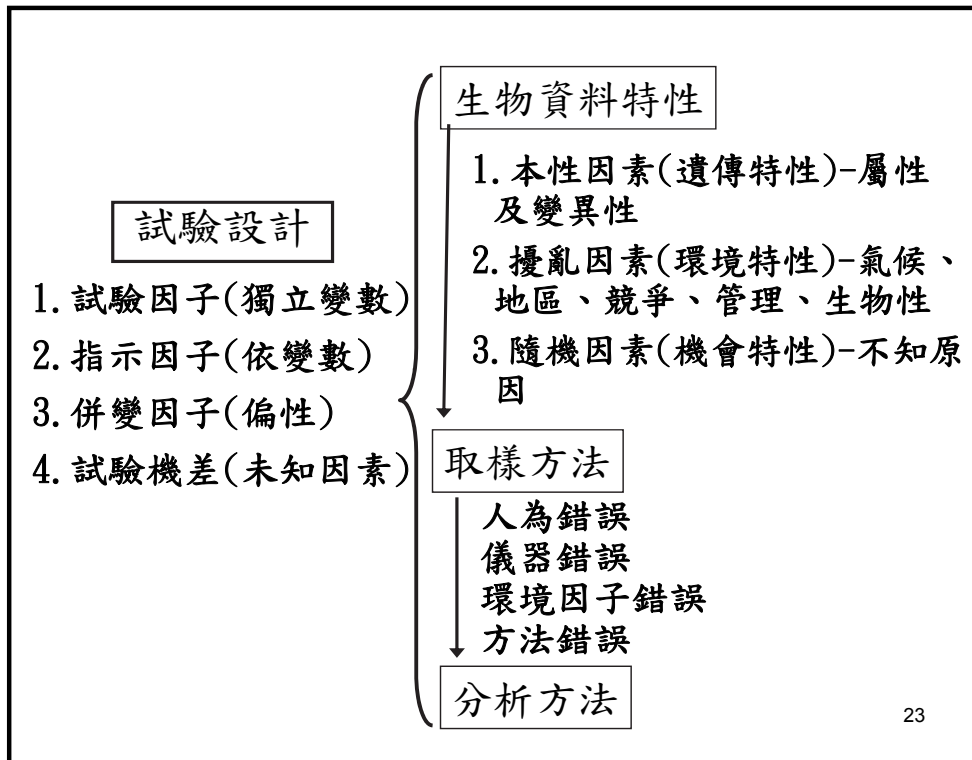
如何設計一個試驗

進行科學試驗的原則

與統計有關

1. 對問題的瞭解及陳述
2. 處理和變級之選擇
3. 反應變數之選擇
4. 試驗設計之選擇(重複、設計方法、取樣方法之決定)
5. 試驗之執行
6. 資料分析
7. 結論及建議

22



試驗設計方法之類別

以參試因子數分類：

單因子試驗、複因子試驗

以隨機方式分類：

完全隨機設計(CRD)、隨機完全區集設計(RCBD)、拉丁方設計(Latin squares design)、裂區設計(Split plot design)、摺疊或巢狀設計(Hierarchical or Nested design)等

以一定順序排列：順序設計

其中以CRD和RCBD最廣受使用

CRD：適於實驗室試驗(材料相當純質且環境偏性易控制)

RCBD：適於田間試驗或有明顯單一偏性因素存在的情況

25

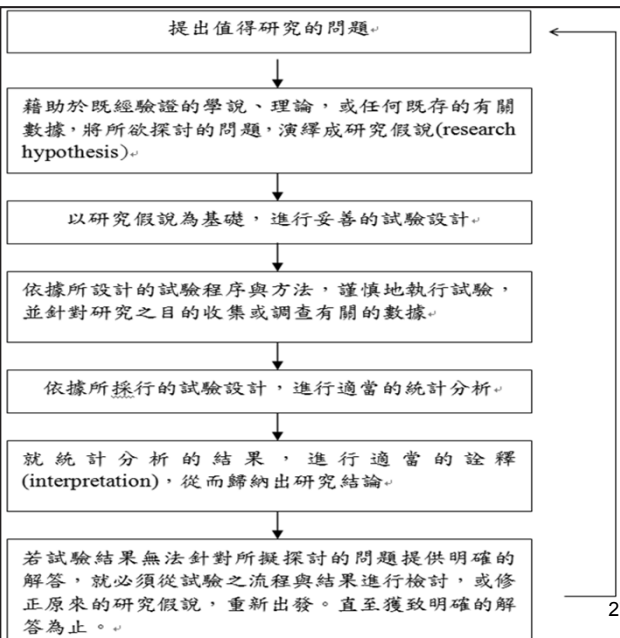
作物改良最常使用的統計方法

- ☞ 一般統計量
- ☞ χ^2 test
- ☞ T test
- ☞ F test
- ☞ 相關分析
- ☞ 回歸分析
- ☞ 多變數統計分析
- ☞ 無母數分析

26

試驗性研究之循環歷程

科學研究是一種尋求真理的過程，而試驗設計的各種原理及方法，其目的皆在儘可能提高這種過程的效率。就試驗性研究而言，由於其結論未必然就是真理，因而形成一種回饋式的循環歷程。



27

作物品種改良與創新研究 —以水稻品種改良為例

☞ 水稻育種過程

親本選定 → 雜交 → 早期世代選拔 → 產量試驗、高級試驗 → 各種特性檢定、肥料試驗 → 區域試驗、米質及食味檢定

☞ 試驗設計方法

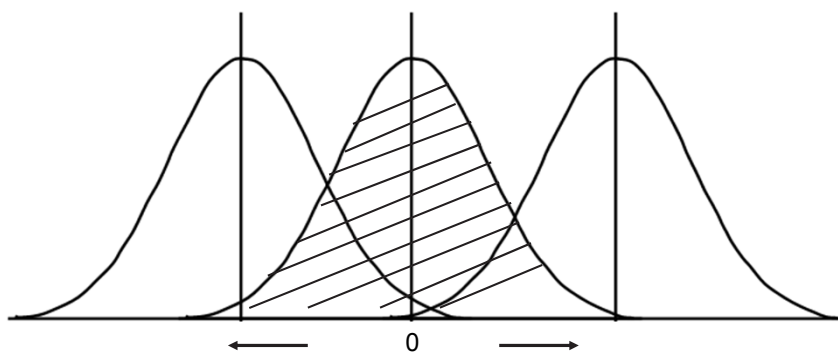
CRD、RCBD、SPD

☞ 統計分析方法

變方分析(ANOVA)、綜合變方分析(CANOVA)、穩定性分析(Stability Analysis)、迴歸分析(Regression Analysis)

28

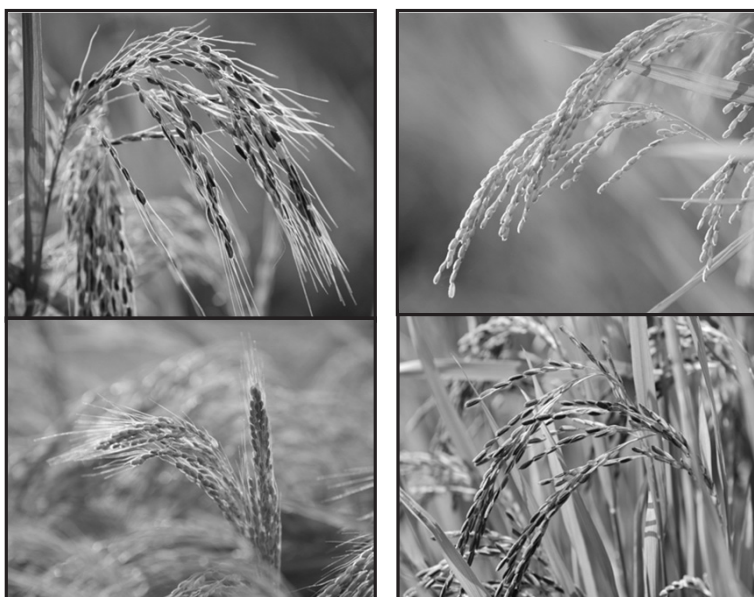
選拔結果—性狀平均值的偏移



透過選拔(以統計方法執行)篩選出符合育種者理想的性狀並留存種子繁衍

29

水稻穀粒與穗性狀的選拔



哪種穀粒型狀、顏色與穗形符合我們的期望，可以透過何種統計方法來協助。

30

單因子RCBD試驗的變方分析與 最小顯著性測驗(LSD)

Source	DF	MS	F
區集	3	1.82	1.93
品種	3	5.48	5.83**
機差	10	0.94	

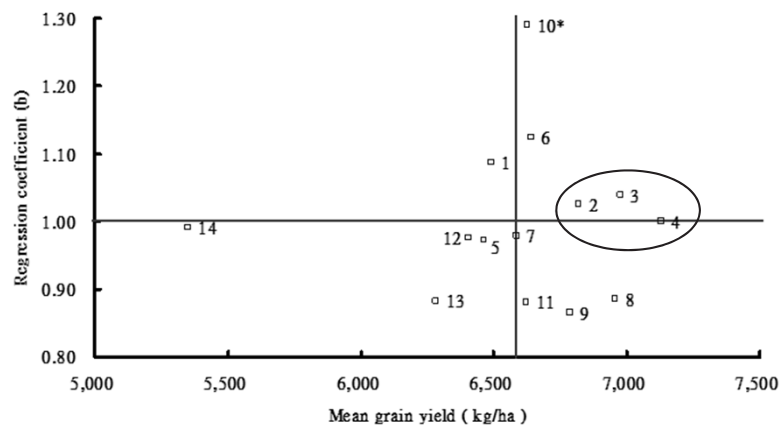
**代表達 1% 顯著性水準

Var.	Mean±SE
C	7250±165 ^{ay}
A	6120±145 ^b
D	6010±210 ^b
B	4530±185 ^c

y: Different letters represent significant treatment difference at 5% level.

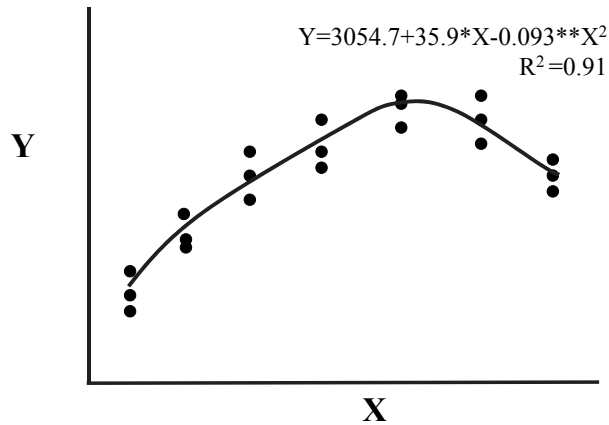
31

稻穀產量穩定性分析的旨在評估新品系在不同環境下之產量表現情形。以 Finlay and Wilkinson (1963) 的公式進行穩定性分析，回歸模式為 $Y_{ij} = a_i + bX_j + e_{ij}$ ，各品系在各地區之產量 Y_{ij} 隨環境效應 X_j 作回歸分析，計算穩定性回歸係數 b ，並以 b 評估各品種的穩定性。



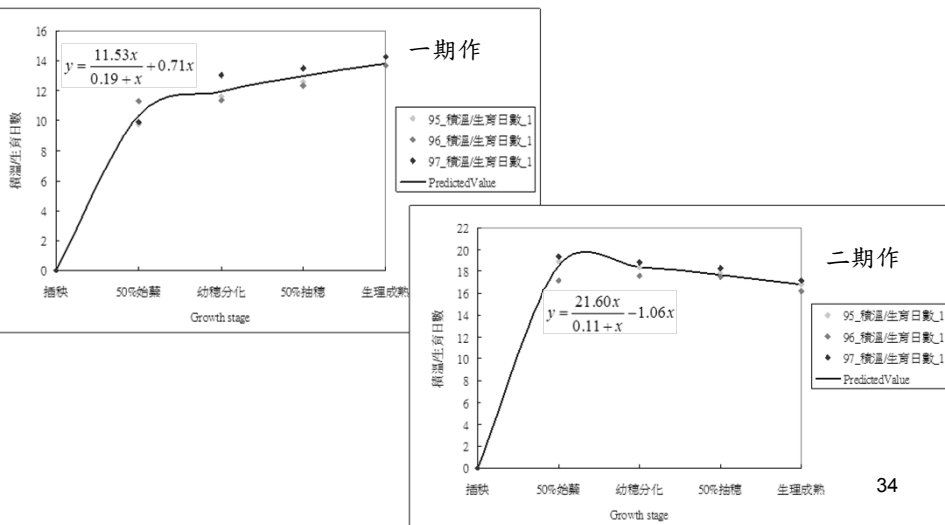
32

某品種水稻在不同氮肥施用量下，氮肥效應對於稻穀產量的反應呈現二次效應，因此當氮肥施用超過某個用量後，其稻穀產量不增反減。



33

水稻台農71號的生長受到溫度的影響，整個生育期的生長曲線呈現非線性多項式，且兩個期作的生長趨勢不同。

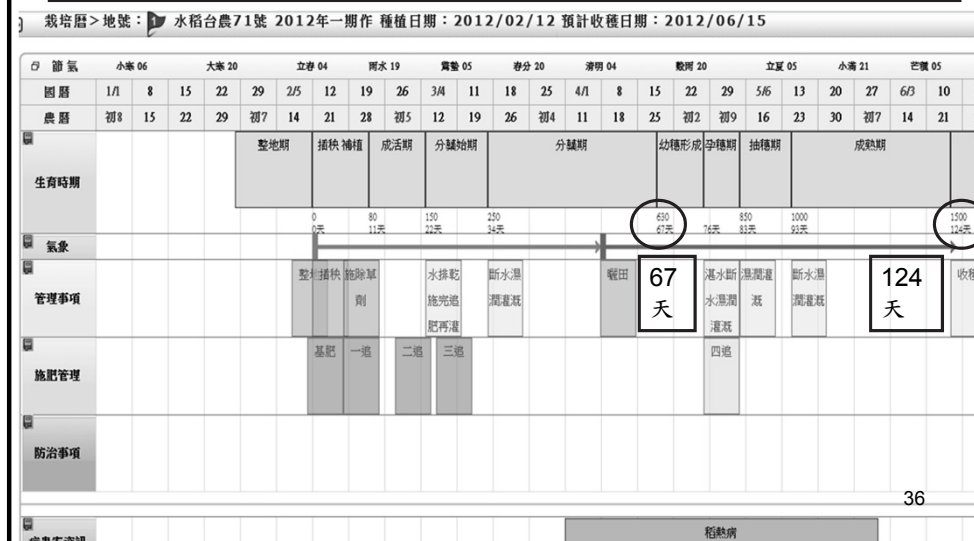


試驗研究與應用

- 例如，肥料試驗與生長發育試驗之成果，透過資訊通訊技術讓水稻的栽培農戶可以免費使用，有助其預測生長期與田間管理之參考。
- 選育不同的植株顏色，透過大地藝術創作，用米食之稻來彩繪大地，注入農村景觀新的泉源。

35

開發生育模式可提供給水稻栽培者用來預估水稻生育時期，作為田間管理的參考指標。





誰知盤中飧 粒粒皆辛苦



38