

# PHƯƠNG PHÁP NGHIÊN CỨU CÂY TRỒNG

# PHẦN 1 - XỬ LÝ SỐ LIỆU ĐIỀU TRA KHẢO SÁT

## Chương 1 - THỐNG KÊ MÔ TẢ - CÁC THAM SỐ THỐNG KÊ

### 1. Tổng thể và mẫu

- Tổng thể
- Mẫu

### 2. Các tham số đặc trưng của mẫu và tổng thể

- Các tham số đặc trưng cho sự tập trung
- Các tham số đặc trưng cho độ phân tán của các dấu hiệu **định lượng**
- Các tham số đặc trưng cho độ phân tán của các dấu hiệu **định tính**
- Các tham số đặc trưng cho mối quan hệ giữa các đại lượng ngẫu nhiên

# 1. Tổng thể và mẫu

- Tổng thể/Dân số : là toàn bộ các phần tử hay cá thể có cùng một hay một số đặc trưng (dấu hiệu) định tính hay định lượng nào đó của đối tượng nghiên cứu

Trong nông học, một tổng thể có thể là:

- Một quần thể cây trồng;
- Một nhân tố cụ thể liên quan đến cây trồng;
- Một khu đất canh;
- ... ..

- Số lượng các phần tử hay cá thể trong tổng thể được ký hiệu là  $N$ .
- Có hai loại tổng thể:
  - \* ***Định tính*** là các dấu hiệu có thể phân biệt sự khác nhau giữa các cá thể bằng mắt, nếm hay thử.
  - \* ***Định lượng*** là các dấu hiệu không thể phân biệt sự khác nhau giữa các cá thể bằng mắt, mà phải tiến hành cân, đo, đếm và phân biệt được nhờ sử dụng các phép toán thống kê.

- **Mẫu**: là một bộ phận hữu hạn của tổng thể gồm  $n$  cá thể ( $n < N$ ), trên đó người ta tiến hành điều tra, khảo sát, đo đếm và thu thập các số liệu.
- Vai trò của mẫu :
  - Ước lượng các tham số của tổng thể.
  - Tìm hiểu mối quan hệ giữa các dấu hiệu nghiên cứu trong tổng thể

- Các phương pháp chọn mẫu
  - + Rút ngẫu nhiên trực tiếp từ tổng thể - có 4 phương pháp chọn ngẫu nhiên
    - *Rút mẫu ngẫu nhiên đơn giản:*
    - *Rút ngẫu nhiên hệ thống:*
    - *Dùng bảng số ngẫu nhiên:*
    - *Dùng phần mềm Excel*
  - + Chọn cá thể điển hình trực tiếp từ tổng thể - Đây là phương pháp chọn mẫu không ngẫu nhiên.
  - + Rút từ các phần của tổng thể (chia nhóm rồi chọn mẫu)

## 2. Các tham số đặc trưng của mẫu và tổng thể

- \* Cực trị - là số bé nhất và lớn nhất trong mẫu, ký hiệu là  $X_{\min}$  và  $X_{\max}$
- \* Mốt (Mode) - là trị số có tần số cao nhất trong một mẫu
- \* Trung bình và kỳ vọng
  - *Trung bình* (trung bình mẫu), ký hiệu là  $\bar{X}$ , là tham số đặc trưng cho sự tập trung của mẫu.
  - *Kỳ vọng* (trung bình tổng thể), ký hiệu là  $E(X)$ ,  $MX$ ,  $\mu$  hay  $m$ , là tham số đặc trưng cho sự tập trung của tổng thể.

# Các tham số đặc trưng cho độ phân tán của các dấu hiệu định lượng

\* Khoảng biến thiên

$$R = X_{\max} - X_{\min}$$

\* Phương sai mẫu, phương sai tổng thể và độ lệch chuẩn



## + Phương sai mẫu và phương sai tổng thể

*Phương sai mẫu* (trung bình bình phương thực nghiệm), ký hiệu là  $S^2$ , là tham số đặc trưng cho độ phân tán của các cá thể trong mẫu.

*Phương sai tổng thể* (trung bình bình phương lý luận), ký hiệu là  $V(X)$ , là tham số đặc trưng cho độ phân tán của các cá thể trong tổng thể.

Ví dụ :

Mau 1	20	24	24	23	25	14	21	20	31	16	18	21	19	20	19	13	20	24	18	20
Mau 2	26	25	29	14	23	13	14	22	28	24	15	31	14	13	29	16	28	14	17	15

Mẫu 1 :

-Trung bình :  $(20 + 24 + \dots + 20) : 20 = 20.5$

-Khoảng biến thiên :  $31 - 13 = 18$

- Phương sai mẫu :  $(20 - 20.5)^2 + (24 - 20.5)^2 + \dots + (20 - 20.5)^2 : (20 - 1) =$   
 $= 16.37$

Mẫu 2 :

-Trung bình :  $(26 + 25 + \dots + 15) : 20 = 20.5$

-Khoảng biến thiên :  $31 - 13 = 18$

- Phương sai mẫu :  $(26 - 20.5)^2 + (25 - 20.5)^2 + \dots + (15 - 20.5)^2 : (20 - 1) =$   
 $= 42.79$

## + Độ lệch chuẩn

*Độ lệch chuẩn mẫu*, ký hiệu là  $S$  hay  $S_d$  (standard deviation)  $S_d = \sqrt{S^2}$

*Độ lệch chuẩn tổng thể*, ký hiệu là  $S_x$  hay  $S$  (nói chung) là căn bậc hai của phương sai tổng thể

$$\sigma = \sqrt{\sigma^2}$$

## + Hệ số biến động

– Hệ số động, ký hiệu là CV (Coefficient of Variation)

$$CV(\%) = \frac{S}{\bar{X}} \cdot 100$$

- Hệ số biến động được sử dụng trong các trường hợp sau:
  - + Đánh giá độ biến động của các cá thể trong mẫu và tổng thể
    - Nếu  $CV(\%) \leq 5\%$  là biến động ít
    - $CV(\%)$  trong khoảng 6 – 10% là biến động vừa phải
    - $CV(\%) > 10$  là biến động nhiều và rất nhiều
  - + Đánh giá sự khác nhau giữa các nhóm cá thể
  - + Chọn ruộng (đất) thí nghiệm.
  - + Đánh giá độ chính xác (ít sai số) của một thí nghiệm

– Hệ số nhọn của phân phối xác suất

$$\alpha_4 = \frac{\mu^4}{\sigma^4}$$

trong đó:

$\mu^4$  là mômen trung tâm bậc 4:

$$\mu^4 = E[X - E(X)]^4$$

$\sigma^4$  là bình phương của phương sai

- Nếu  $\alpha_4 = 3$  thì đồ thị phân phối xác suất là bình thường
- Nếu  $\alpha_4 > 3$  thì đồ thị nhọn (các xi tập trung nhiều xung quanh kỳ vọng  $m$ )
- Nếu  $\alpha_4 < 3$  thì đồ thị tù (không nhọn).

# Các tham số đặc trưng cho độ phân tán của các dấu hiệu định tính

+ Độ lệch chuẩn

+ Hệ số biến động

Độ lệch chuẩn được tính :

$$S_p = \sqrt[k]{p_1 \cdot p_2 \cdots p_k}$$

Hệ số biến động được tính

$$CV(\%) = \frac{S_p}{S_{p\max}} \cdot 100$$

Theo mức độ lông của lá, Bảng 1.1 cho thấy trong 28 giống được khảo sát có

7,1% số giống không hay rất ít lông;

14,3% - ít lông;

21,4% - lông vừa;

28,6% - lông nhiều;

28,6% - lông rất nhiều.

Trong 20 giống kháng rầy có

- 5% thuộc loại ít lông;

- 20% thuộc loại lông vừa;

- 35% thuộc loại nhiều lông;

- 40% thuộc loại rất nhiều lông

Đánh giá độ lệch chuẩn

**Bảng 1.1:** Kết quả điều tra mức độ rầy xanh hại trên 28 giống bông tại Đại học Nông Lâm Tp. HCM, 2009

Mức độ lông của lá	Giống có lông		Giống kháng	
	Số giống	Tần suất	Số giống kháng	Tần suất
	$(n_i)$	$(p_i)$	$(n_i)$	$(p_i)$
Không hay rất ít	2	0,071	0	0,000
Ít	4	0,143	1	0,050
Vừa	6	0,214	3	0,150
Nhiều	8	0,286	7	0,350
Rất nhiều	8	0,286	8	0,400
<b>Tổng</b>	<b>28</b>	<b>1,000</b>	<b>20</b>	<b>1,000</b>



Về giống có lông

$$S_p = \sqrt[5]{0,071 \times 0,143 \times 0,214 \times 0,286 \times 0,286}$$

$$= 0.178 \text{ (17.8\%)}$$

$$CV(\%) = \frac{0,178}{0,200} \cdot 100$$

$$= 89,0$$

Về giống có kháng : một cách tương tự ta có:

$$S_p = 0,193 \text{ và}$$

$$CV(\%) = 77,4$$

- **Hiệp phương sai**

Hiệp phương sai, ký hiệu là  $\text{Cov}(X, Y)$ , là kỳ vọng của tích các độ lệch của các đại lượng ngẫu nhiên với kỳ vọng của chúng, biểu thị mức độ quan hệ giữa hai đại lượng ngẫu nhiên

$$W_{x,y} = \frac{1}{n-1} \sum (X - \bar{X})(Y - \bar{Y})$$

Hoặc

$$\text{Cov}(X, Y) = \frac{1}{n-1} \left[ \sum XY - (\sum X)(\sum Y) / n \right]$$