

ỦY BAN NHÂN DÂN TỈNH ĐỒNG THÁP
TRƯỜNG CAO ĐẲNG CỘNG ĐỒNG ĐỒNG THÁP



GIÁO TRÌNH

MÔN HỌC: THỐNG KÊ VÀ PHƯƠNG PHÁP

THÍ NGHIỆM - KHCT

NGÀNH, NGHỀ: KHOA HỌC CÂY TRỒNG

TRÌNH ĐỘ: CAO ĐẲNG

*(Ban hành kèm theo Quyết định Số: .../QĐ-CD-ĐT ngày... tháng... năm
2017 của Hiệu trưởng Trường Cao đẳng Cộng đồng Đồng Tháp)*

Đồng Tháp, năm 2017

TUYÊN BỐ BẢN QUYỀN

Tài liệu này thuộc loại sách giáo trình nên các nguồn thông tin có thể được phép dùng nguyên bản hoặc trích dùng cho các mục đích về đào tạo và tham khảo.

Mọi mục đích khác mang tính lệch lạc hoặc sử dụng với mục đích kinh doanh thiếu lành mạnh sẽ bị nghiêm cấm.

TaiLieu.vn

LỜI GIỚI THIỆU

Thống kê phép thí nghiệm là một môn khoa học thông dụng được giảng dạy trong một số trường đại học, cao đẳng có liên quan đến lĩnh vực Nông Nghiệp như KHOA HỌC CÂY TRỒNG, Khoa Học Cây Trồng, Thủy Sản, Chăn Nuôi... nhằm phục vụ cho công tác thu thập số liệu, xử lý số liệu, bố trí thí nghiệm.

Giáo trình này được viết cho sinh viên bậc cao đẳng ngành, nghề KHOA HỌC CÂY TRỒNG nên giáo trình chỉ trình bày các nội dung thống kê cơ bản mang tính ứng dụng như khái niệm cơ bản về thống kê, cách bố trí thí nghiệm ngoài đồng và cách sử dụng chương trình Excel phần mềm MstatC để kiểm định giá trị trung bình của các nghiệm thức trong bố trí thí nghiệm 1 và 2 nhân tố.

Nội dung tài liệu gồm 5 chương

Chương 1. Một vài khái niệm thường dùng trong thống kê

Chương 2: Các dạng phân bố của biến ngẫu nhiên

Chương 3. So sánh hai mẫu độc lập

Chương 4: Bố trí thí nghiệm

Chương 5: Phân tích kết quả thí nghiệm một nhân tố

Chương 6: Phân tích kết quả thí nghiệm hai nhân tố

Giáo trình được biên soạn nhằm phục vụ cho việc giảng dạy trình độ cao đẳng ngành, nghề KHOA HỌC CÂY TRỒNG tại trường CĐCD Đồng Tháp. Trong quá trình biên soạn không tránh khỏi nhiều thiếu sót. Rất mong nhận được sự đóng góp ý kiến quý báu của anh chị em đồng nghiệp và bạn đọc để chúng tôi bổ sung, chỉnh sửa cho giáo trình ngày càng hoàn thiện, góp phần vào sự nghiệp đào tạo nghề KHOA HỌC CÂY TRỒNG trong tỉnh được tốt hơn.

Xin bày tỏ lòng biết ơn với Lãnh đạo trường CĐCD Đồng Tháp, Hội Đồng thẩm định đã đóng góp nhiều ý kiến quý báu để hoàn chỉnh giáo trình. Cảm ơn các tác giả biên soạn những tài liệu tôi tham khảo và bạn bè đồng nghiệp đã giúp đỡ, cung cấp nhiều tài liệu để tôi hoàn thành giáo trình này.

Đồng Tháp, ngày.....tháng ... năm 2017

Chủ biên

Nguyễn Thị Quế Phương

MỤC LỤC

Trang

Table of Contents

LỜI GIỚI THIỆU	ii
CHƯƠNG 1	1
1. Một số khái niệm.....	1
1.1. Thống kê và thống kê sinh học	1
1.2. Tổng thể và mẫu.....	2
1.3. Các loại biến số	9
1.4. Các loại thang đo trong thống kê	10
2. Dữ liệu.....	11
2.1. Khái niệm.....	11
2.2. Các loại dữ liệu	12
2.3. Các phương pháp thu thập dữ liệu.....	12
2.4. Mô tả và trình bày dữ liệu.....	14
3. Thực hành.....	18
3.1. Tính toán các tham số mô tả bộ dữ liệu	18
3.2. Mô tả và trình bày dữ liệu.....	19
CHƯƠNG 3	41
KIỂM ĐỊNH GIẢ THIẾT SỬ DỤNG 2 MẪU	41
1. Nguyên tắc	41
1.1. So sánh trung bình hai mẫu.....	41
1.2. So sánh cặp.....	44
1.3. So sánh tỉ lệ hai mẫu	45
2. Thực hành.....	45
2.1. T-test	45
2.2. Z- tesst.....	52

CHƯƠNG 4	57
THIẾT KẾ THÍ NGHIỆM	57
1. Một số định nghĩa thường dùng trong bố trí thí nghiệm.....	57
1.1. Đơn vị thí nghiệm (Experimental unit).....	57
1.2 Nhân tố (Factor)	58
1.3. Nghiệm thức (treatment)	58
1.4. Sai số thí nghiệm (Experimental error).....	59
1.5. Lặp lại (Replication)	60
2. Thiết kế thí nghiệm một nhân tố	60
2.1. Bố trí hoàn toàn ngẫu nhiên	61
2.2. Bố trí khối hoàn toàn ngẫu nhiên	64
2.3. Bố trí thí nghiệm theo kiểu hình vuông Latin.....	67
3. Thiết kế thí nghiệm hai nhân tố	70
3.1. Bố trí thí nghiệm kiểu khối hoàn toàn ngẫu nhiên.....	71
3.2. Bố trí thí nghiệm theo kiểu thừa số lô phụ	72
4. Thực hành: Các kiểu bố trí thí nghiệm	74
CHƯƠNG 5	76
PHÂN TÍCH KẾT QUẢ THÍ NGHIỆM MỘT NHÂN TỐ	76
1. Bố trí hoàn toàn ngẫu nhiên	76
1.1. Phân tích phương sai	76
Bảng 4.1. Phân tích phương sai (CRD)	77
1.2. Kiểm định sự khác biệt	79
2. Bố trí khối hoàn toàn ngẫu nhiên	82
2.1. Phân tích phương sai	82
2.2. Kiểm định sự khác biệt	84
3. Thực hành.....	84
3.1. Phương pháp nhập số liệu	84
3.2. Bố trí hoàn toàn ngẫu nhiên một nhân tố.....	87
3.3. Bố trí khối hoàn toàn ngẫu nhiên một nhân tố.....	92

CHƯƠNG 6	107
PHÂN TÍCH KẾT QUẢ THÍ NGHIỆM HAI NHÂN TỔ	107

TaiLieu.vn

GIÁO TRÌNH MÔN HỌC

Tên môn học: THỐNG KÊ VÀ PHƯƠNG PHÁP THÍ NGHIỆM- KHCT

Mã môn học: NN206

Vị trí, tính chất, ý nghĩa và vai trò của môn học/môn học:

- Vị trí: Môn học thống kê phép thí nghiệm được bố trí sau môn Tin học, trước môn học Thực tập tốt nghiệp
- Tính chất: Là môn học cơ sở, hướng dẫn cách bố trí thí nghiệm, sắp xếp, xử lý số liệu, đọc kết quả thống kê, làm nền tảng cho mô đun Thực tập tốt nghiệp
- Ý nghĩa và vai trò của môn học: *giúp cho sinh viên tiếp cận môn học dễ dàng, hiểu được các ứng dụng thực tế của môn học trong thực tập cuối khóa và nghiên cứu khoa học.*

Mục tiêu của môn học/môn học:

- Về kiến thức:
 - + Phát biểu được các khái niệm dùng trong thống kê;
 - + Giải thích được số liệu đã qua xử lý thống kê của thí nghiệm 2 mẫu độc lập;
 - + Phát biểu được các khái niệm dùng trong bố trí thí nghiệm 1 và 2 nhân tố
 - + Phát biểu được các phương pháp bố trí thí nghiệm một nhân tố và 2 nhân tố.
 - + Giải thích được số liệu đã qua xử lý thống kê thí nghiệm 1 nhân tố
- Về kỹ năng:
 - + Sắp xếp các số liệu theo nhóm, lớp
 - + Tính toán được các số đo mô tả
 - + Trình bày dữ liệu bằng excel
 - + Sử dụng phần mềm thống kê để xử lý số liệu từ kết quả thí nghiệm
 - + Chọn lựa và thực hiện được công tác bố trí thí nghiệm để thu thập số liệu cho công tác nghiên cứu;
 - + Tính toán được các số đo mô tả, bảng phân tích phương sai và kiểm định sự khác biệt giữa các nghiệm thức;

- + Trình bày kết quả thống kê
- Về năng lực tự chủ và trách nhiệm:
 - + Rèn luyện tính cẩn thận, chính xác, ham học hỏi. có thể tự xác định các chỉ tiêu thu thập, đánh giá kết quả thí nghiệm và đưa ra nhận định cho kết quả đã phân tích

Nội dung của môn học:

Số TT	Tên chương, mục	Thời gian (giờ)			
		Tổng số	Lý thuyết	Thực hành, thí nghiệm, thảo luận, bài tập	Kiểm tra
	Chương 1: Một vài khái niệm thường dùng trong thống kê 1 Tập hợp 2 Mẫu và cỡ mẫu 3 Biến ngẫu nhiên 4 Các số đo mô tả	4	4	0	
	Chương 2: Các dạng phân bố của biến ngẫu nhiên 1 Phân bố nhị thức 2 Phân bố chuẩn 3 Phân bố của trung bình mẫu 4 Phân bố của số tỉ lệ mẫu 5 Phân bố Student 6 Phân bố Fisher	4	4	0	
3	Chương 3: So sánh hai mẫu độc lập 1. T-test 2. Z-test 3. Tương quan, hồi qui 4. Thực hành	5	1	4	
4	Chương 4: Thiết kế thí nghiệm	8	3	4	1LT

	1 Một số định nghĩa thường dùng trong bố trí thí nghiệm 2 Thiết kế thí nghiệm một nhân tố 3. Thiết kế thí nghiệm hai nhân tố 4. Thực hành				
5	Chương 5: Phân tích kết quả thí nghiệm một nhân tố 1. Bố trí hoàn toàn ngẫu nhiên 2. Bố trí khối hoàn toàn ngẫu nhiên 3. Thực hành	12	4	8	
6	Chương 6: Phân tích kết quả thí nghiệm hai nhân tố 1. Bố trí khối hoàn toàn ngẫu nhiên 2. Bố trí lô phụ 3. Thực hành	7	3	3	1TH
	Cộng	40	19	19	2

CHƯƠNG 1

MỘT VÀI KHÁI NIỆM THƯỜNG DÙNG TRONG THỐNG KÊ

NN206-01

Giới thiệu

Chương học trình bày các khái niệm dùng trong thống kê, phương pháp thu thập, mô tả và trình bày dữ liệu thống kê

Mục tiêu:

Kiến thức:

- + Phát biểu được các khái niệm dùng trong thống kê

Kỹ năng:

- + Sắp xếp các số liệu theo nhóm, lớp
- + Tính toán được các số đo mô tả
- + Trình bày dữ liệu bằng excel

Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, chính xác, ham học hỏi. Quyết định phương pháp trình bày dữ liệu phù hợp tình huống cụ thể

1. Một số khái niệm

1.1. Thống kê và thống kê sinh học

Thuật ngữ thống kê có hai nghĩa: Nghĩa thứ nhất, thống kê là những con số được ghi chép để phản ánh các hiện tượng của tự nhiên (lượng mưa, nhiệt độ), kỹ thuật, kinh tế, xã hội (dân số, lao động)... Theo nghĩa thứ hai, thống kê là hệ thống các phương pháp thu thập và phân tích các con về những hiện tượng nói trên để tìm hiểu bản chất và tính quy luật vốn có của nó. Chẳng hạn qua số liệu về chiều cao, đường kính, năm tuổi, độ che phủ, tỉ lệ dịch hại, năng suất... của một loại cây trồng, ta có thể tìm hiểu được mức độ quan hệ giữa các đại lượng, dự báo chiều cao hoặc độ che phủ của cây qua năm tuổi và đường kính của nó,... từ đó giúp người quản lý quyết định về mật độ trồng, kế hoạch chăm sóc, khai thác, một cách hợp lý. Trong giáo trình này, phần thống kê chủ yếu trình bày các vấn đề theo nghĩa thứ hai.

Thống kê sinh học bắt nguồn từ tiếng Hy Lạp bios sự sống và metron đo đạc nên người ta gọi đây là sinh trắc (biological measurement). Thống kê sinh học là khoa học về sự ứng dụng các phương pháp thống kê để giải quyết các vấn đề trong sinh học.

Để thực hiện một thống kê đầy đủ, thông thường thực hiện hai loại sau:

- Thống kê mô tả: Thu thập và kiểm tra số liệu, mô tả và trình bày số liệu, tính các tham số mẫu đặc trưng của số liệu mẫu.

- Thống kê suy diễn: Thực hiện việc ước lượng, kiểm định, phân tích mối liên hệ, dự đoán,... trên cơ sở các thông tin thu thập từ mẫu.

1.2. Tổng thể và mẫu

1.2.1. Tổng thể (Population)

Tổng thể hay còn gọi là tập hợp bao gồm tất cả các phần tử (đơn vị) thuộc đối tượng nghiên cứu. Giá trị một phần tử trong tập hợp gọi là biến số. Mỗi phần tử trong tập hợp gọi là cá thể. Số cá thể của tổng thể được ký hiệu là N .

Tổng thể có thể là hữu hạn hoặc vô hạn (infinite)

Ví dụ: tập hợp chiều cao của tất cả sinh viên trong lớp học là tổng thể hữu hạn vì có thể xác định được số sinh viên trong lớp học là bao nhiêu. Nhưng tập hợp chiều cao của giống lúa IR50404 trong ruộng thí nghiệm là tổng thể vô hạn vì không thể đo chiều cao tất cả cây trong ruộng được.

Một ví dụ khác như tập hợp của hàm lượng Vitamin C của các trái quýt hồng chín trong vườn cây là tổng thể vô hạn vì số trái quýt hồng chín trong vườn rất nhiều, ta không thể định lượng hàm lượng Vitamin C một cách chính xác hết tất cả các trái nên trường hợp này được xem là tổng thể vô hạn.

1.2.2. Mẫu (Sample)

Nhìn chung trong lĩnh vực nghiên cứu khoa học thuộc ngành nông nghiệp, tổng thể thường là vô hạn, chúng ta không thể nào quan sát hết tất cả các cá thể trong tổng thể mà chỉ có thể quan sát một nhóm cá thể được rút ra từ tổng thể gọi là mẫu.

Mẫu là một bộ phận hay một số cá thể (phần tử) được rút ra từ tập hợp. Số phần tử chứa trong mẫu gọi là cỡ mẫu hay kích thước mẫu (Sample size). Ký hiệu cỡ mẫu là n

a. Phương pháp chọn mẫu

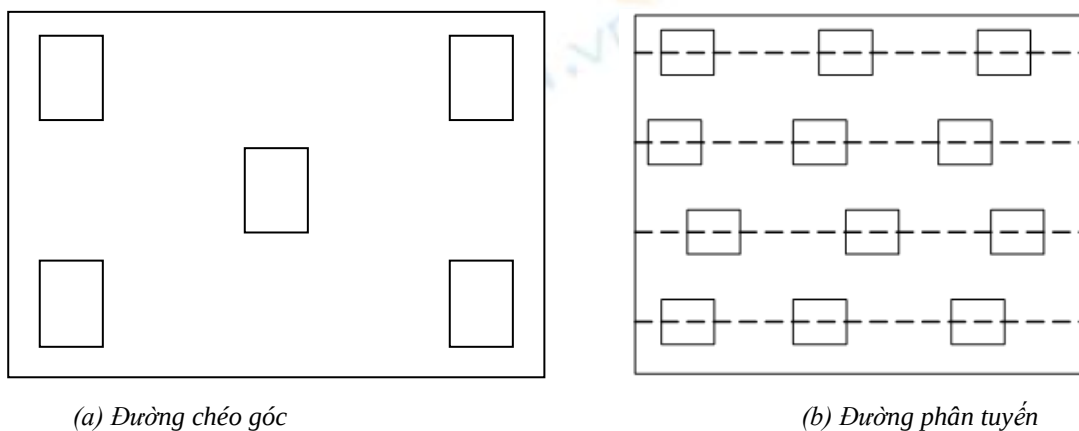
Mẫu phải đảm bảo tính đại diện, khách quan, chính xác và dựa trên quan điểm toán học xác suất thống kê. Tuy nhiên, để đạt mục đích trên còn phải kết hợp với cả độ lớn của mẫu nghiên cứu mới đầy đủ.

- Chọn mẫu ngẫu nhiên: đây là phương pháp chọn mẫu mà các cá thể được lấy ra quan sát, đo đếm là hoàn toàn ngẫu nhiên. Cách thực hiện: toàn bộ các cá thể trong ô thí nghiệm được đánh số sau đó bốc thăm hoặc tra bảng ngẫu nhiên (Phụ lục 1) để chọn ra được các cá thể của mẫu (loại trừ các cá thể ở hàng biên).

+ Ưu điểm: mẫu nghiên cứu mang tính khách quan và các giá trị thu được tuân theo quy định của đại lượng ngẫu nhiên, do đó các tham số của mẫu mang tính đại diện, nhưng các cá thể trong mẫu mang tính biến động (không đồng đều). Song đó là hiện trạng của thí nghiệm (tính chân thực) độ chính xác của kết quả nghiên cứu cao.

+ Nhược điểm: khi số lượng mẫu (cỡ mẫu) không đủ lớn có thể dẫn đến kết quả không chính xác (tính đại diện thấp). Bên cạnh đó việc thực hiện lấy mẫu phức tạp và tốn thời gian.

- Chọn mẫu phân phối đều: Chọn phân phối đều ở đây có thể thực hiện trên ô thí nghiệm, hoặc trên khu vực điều tra. Phân phối đều có 2 dạng: đường chéo góc và đường phân tuyến



Hình 1.1 Cách chọn mẫu phân phối đều.

b. Tham số đặc trưng của mẫu

*** Các số đo trung tâm**

- **Số trung bình cộng** (*arithmetical mean*) được tính bằng tổng các giá trị quan sát trong bộ số liệu chia cho tổng số phần tử quan sát là giá trị ở giữa, số trung bình thường được sử dụng làm giá trị đại diện cho bộ số liệu

- Trung bình của tập hợp: ký hiệu là μ

Công thức tính giá trị trung bình:

$$\mu = \frac{X_1 + X_2 + \dots + X_N}{N} = \sum_{i=1}^N X_i / N$$

Thường trong thực tế chúng ta không biết được μ do kích thước của tập hợp thường quá lớn mà chúng ta chỉ có thể ước lượng được μ bằng $\bar{x}$ (trung bình của mẫu). Do đó nếu n cá thể của mẫu được rút ra từ tập hợp (tổng thể) thì trị số trung bình của mẫu là

$$\hat{\mu} = \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

Ví dụ: Chiều cao (cm) của giống lúa IR50404 vào thời điểm 20 ngày sau khi sạ là: 15, 21, 20, 19, 22, 21, 16, 19, 20, 16, 17, 24, 16, 21, 15 và 22.

$$\hat{\mu} = \bar{X} = \frac{304}{16} = 19 \text{ cm}$$

Nếu số liệu được trình bày theo bảng phân bố thực nghiệm, trung bình có thể được tính theo hai cách sau:

- Dựa trên tần số

$$\bar{X} = \frac{\sum_{i=1}^N X_i f_i}{n}$$

X_i : là giá trị của biến quan sát thứ i
 f_i : là tần số của giá trị X_i

- Dựa trên tần suất:

$$\bar{X} = \sum_{i=1}^N X_i F_i$$

X_i là giá trị của biến quan sát thứ i
 f_i là tần suất của giá trị X_i

Bảng 1.1. Bảng phân bố thực nghiệm

Chiều cao (cm)	Tần số (f)	Tần suất (F)	fX	FX
15	2	0.1250	30	1.8750
16	3	0.1875	48	3.0000
17	1	0.0625	17	1.0625
19	2	0.1250	38	2.3750
20	2	0.1250	40	2.5000
21	3	0.1875	63	3.9375
22	2	0.1250	44	2.7500
24	1	0.0625	24	1.5000
	16	1	304	19

$$\Rightarrow \bar{X} = \frac{\sum_{i=1}^n X_i f_i}{n} = 304/16 = 19 \text{ cm} \quad \text{hoặc} \quad \bar{X} = \sum_{i=1}^n X_i F_i = 19 \text{ cm}$$

- **Số trung vị (Median) – Me:** Là trị số giữa của một chuỗi số liệu đã được sắp xếp thứ tự

Cách xác định số trung vị:

Khi bộ số liệu là các giá trị rời rạc

- Nếu n là số lẻ, trung vị là số thứ $(n+1)/2$

Ví dụ: 3 7 9 10 12

Có $n = 5$ vậy vị trí số trung vị $= (5+1)/2 = 3$

Số trung vị ở vị trí thứ 3 trong bộ số liệu là số 9 (không phải là số trung bình)

- Nếu n là số chẵn: trung vị là giá trị trung bình của 2 trị số thứ $(n/2)$ và $(n/2)+1$

Ví dụ: 1 3 7 8 12 16

$Me = (7+8)/2 = 7.5$ nghĩa là số trung bình của số hạng thứ 3 và số hạng thứ 4

Số trung vị là số đại diện cho một nhóm số, nó có tính trội hơn số trung bình là không bị ảnh hưởng bởi một vài số liệu quá lớn hay quá nhỏ ở một đầu của chuỗi số liệu. Do đó, trong một số trường hợp số trung vị được xem là một giá trị trung tâm điển hình hơn cho chuỗi số liệu so với số trung bình. Tuy nhiên, trong trường hợp số liệu được phân nhóm thì cách xác định số trung vị đôi khi dài dòng.

* Các số đo phân tán

- Phương sai (variance)

Là tham số đặc trưng tiêu biểu nhất cho tính chất phân tán của tổng thể

- Phương sai của tổng thể là trung bình độ lệch bình phương của các giá trị quan sát (X_i) so với giá trị trung bình cộng (μ), ký hiệu σ^2

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad N: \text{kích thước của tổng thể}$$

- Phương sai của mẫu. Thường trong thực tế chúng ta không biết được trị số thật của σ^2 , vì không biết μ , mà chỉ ước lượng σ^2 bằng cách dùng số thống kê từ một mẫu ngẫu nhiên, gọi là phương sai mẫu.

Giả sử có n cá thể

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (1)$$

$\bar{X}$: ước lượng của tập hợp (trung bình mẫu)

$$s^2 = \frac{\sum_{i=1}^n X_i^2 - [(\sum_{i=1}^n X_i)^2 / n]}{n - 1}$$

Công thức:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - [(\sum_{i=1}^n X_i)^2 / n]$$

Đây là công thức tính tổng bình phương các độ lệch gọi tắt là tổng bình phương (sum of squares) ký hiệu SS

* n-1: độ tự do (degree of freedom = df)

* $\sum_{i=1}^n X_i^2 / n$: yếu tố hiệu chỉnh (correction factor = C.F.)

Đối với các số liệu được viết dưới dạng tần số

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 f_i}{n - 1}$$

$$s^2 = \frac{\sum_{i=1}^n X_i^2 f_i - [(\sum_{i=1}^n X_i f_i)^2 / n]}{n - 1}$$

Ví dụ: Lấy lại chiều cao cây lúa trong ví dụ trước

X	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X)^2$
15	- 4	16	225
21	2	4	441
20	1	1	400
19	0	0	361
22	3	9	484
21	2	4	441
16	- 3	9	256
19	0	0	361
20	1	1	400
16	- 3	9	256
17	- 2	4	289
24	5	25	576
16	- 3	9	256
21	2	4	441
15	- 4	16	225
22	3	9	484
304	0	120	5896

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} = \frac{120}{15}$$

$$\text{hoặc } s^2 = \frac{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}{n - 1} = \frac{5896 - \frac{304^2}{16}}{15} = 8$$

- **Độ lệch chuẩn (Standard deviation):** là căn bậc hai của phương sai. Nhà toán học người Nga P.L. Chebychev (1821 - 1894) đã dùng số đo độ lệch chuẩn để đo lường độ phân tán của tổng thể. Số đo này có ưu điểm là cùng đơn vị với số liệu đo trong khi phương sai là đơn vị bình phương

$$\sigma = \sqrt{\sigma^2} \quad (\text{tổng thể})$$

$$s = \sqrt{s^2} \quad (\text{mẫu})$$

- **Hệ số biến thiên = hệ số biến động (Coefficient of Variation):** $c.v\%$

Hệ số biến động cũng được dùng để đo lường độ phân tán của tổng thể. Đó là phần trăm tỉ số giữa độ lệch chuẩn và trung bình

$$c.v.(%) = \frac{\sigma}{\mu} \cdot 100$$

Vì σ và μ không biết nên hệ số này được ước lượng bằng hệ số biến thiên của mẫu

$$c.v..(%) = \frac{s}{\bar{X}} \cdot 100$$

Với số liệu Vit.C, $c.v.$ được tính như sau:

$$c.v..(%) = \frac{2,8284}{19} \cdot 100 = 14,89\%$$

Số đo độ lệch chuẩn là một trị số tuyệt đối có cùng đơn vị với trị số trung bình. Tuy nhiên, số đo này không thể dùng để so sánh mức độ biến động của hai hay nhiều chuỗi số liệu có đơn vị đo lường khác nhau. Để khắc phục tình trạng biến động của đơn vị, Pearson đã đưa ra khái niệm hệ số biến thiên (cv). Đây là một trị số tương đối chỉ độ chính xác của việc so sánh các giá trị trung bình và là chỉ số cho phép đánh giá mức độ tin cậy của thí nghiệm. Giá trị của $c.v.$ càng cao thì độ tin cậy của thí nghiệm càng thấp.

Hệ số biến động có ưu điểm hơn độ lệch chuẩn ở chỗ giúp ta so sánh độ phân tán của hai tập hợp số liệu có số trung bình khác nhau và có đơn vị đo lường khác nhau.

Ví dụ : Số liệu về chiều cao và trọng lượng 1000 hạt của 100 mẫu lúa như sau:

* Chiều cao $\bar{X} = 110$ cm, $\sigma = 10$ và $cv = 10\%$

* Trọng lượng $\bar{X} = 35$, $\sigma = 5$ và $cv = 20\%$

Như vậy, số liệu về trọng lượng phân tán hơn số liệu về chiều cao

Giá trị $c.v.$ thay đổi theo từng kiểu thí nghiệm, cây trồng và tính trạng đo lường. Tuy nhiên, một nghiên cứu viên có kinh nghiệm có thể quyết định một cách hợp lý việc chấp nhận giá trị của $c.v.$ cho kiểu thí nghiệm đang khảo sát. Ví dụ, đối với tính trạng năng suất lúa của các thí nghiệm ngoài đồng, khoảng $c.v.$ có

thể chấp nhận là 6 - 8% cho các thí nghiệm về giống; 10 - 12% cho các thí nghiệm về phân bón và 13 - 15% cho các thí nghiệm về thuốc trừ sâu và thuốc trừ cỏ.

Giá trị c.v. của các tính trạng khác thường khác với c.v. của năng suất. Ví dụ, với các thí nghiệm ngoài đồng thì c.v. của năng suất lúa khoảng 10%, c.v. của số chồi khoảng 20% và c.v. của chiều cao cây khoảng 3%.

Bằng cách so sánh c.v., chúng ta có thể đánh giá mức độ chính xác trong việc tiến hành thí nghiệm. Ngoài ra, vì tính chất ổn định của c.v. nên trong nhiều trường hợp có thể dùng c.v. để ước lượng độ lệch chuẩn (s) và xác định cỡ mẫu (n) để thu thập.

1.3. Các loại biến số

Biến ngẫu nhiên là một đại lượng bằng số mà giá trị của nó tùy thuộc vào cách lấy ngẫu nhiên do cân, đong, đo, đếm, quan sát... có được.

Gọi X_i là giá trị của biến X trong quan sát lần thứ i . Nếu có n biến quan sát thì giá trị của các biến lần lượt là $X_1, X_2, \dots, X_n$.

Ví dụ: độ ngọt của dưa hấu lúc chín được xác định thông qua độ brix. Đây là một đại lượng ngẫu nhiên quan sát trên nhiều trái dưa hấu khác nhau thu trên ruộng. Mỗi một độ brix đo được trên một trái dưa hấu được xem là một giá trị của biến ngẫu nhiên X , đo độ brix của bao nhiêu cây mía sẽ thu được bấy nhiêu giá trị X .

Có hai loại biến số:

1.3.1 Biến số định tính (số liệu thuộc tính = qualitative data)

Là dữ liệu đối với thông tin chỉ tính chất của đối tượng được khảo sát, dữ liệu này không dùng các phép tính để tính toán được như giống cây trồng, màu hoa, màu sắc thịt trái, những loại số liệu rất khó có khả năng định lượng chính xác, trong trường hợp này người ta định ra các tiêu chuẩn, trên cơ sở đó sẽ sắp xếp các số liệu thu thập được vào nhóm như cấp độ nhiễm sâu, bệnh hại: rất nặng, nặng, trung bình, nhẹ, cấp bệnh 1, 3, 5...

1.3.2 Biến số liệu định lượng (quantitative data)

Là số liệu có được thông qua việc đo lường hay tính toán như: chiều cao cây, trọng lượng trái, năng suất,... Biến số định lượng cũng được phân thành hai loại

+ Biến ngẫu nhiên rời rạc (discrete random variable): Là biến ngẫu nhiên có giá trị là những con số nguyên thông qua tác động đếm

Ví dụ: - Số hạt lúa trên bông lúa, số chồi của một bụi lúa

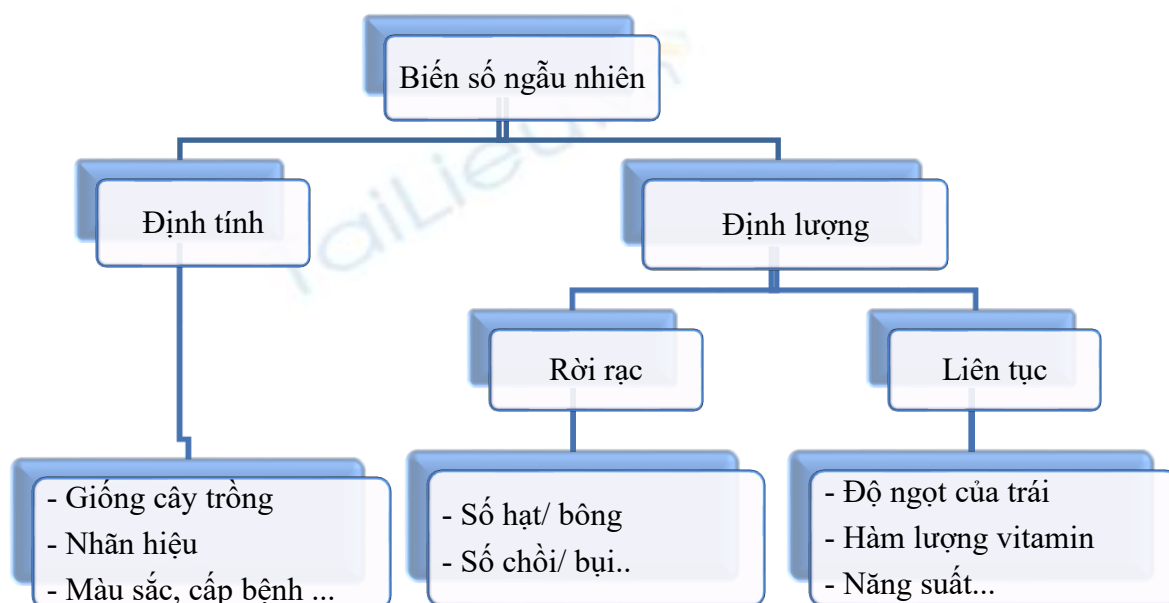
- Số hạt sen trên một gương sen

+ Biến ngẫu nhiên liên tục (continuous random variable): Là biến ngẫu nhiên mà các giá trị của nó có được thông qua đo lường hay tính toán. Biến liên tục có thể nhận giá trị bất kỳ trong một khoảng các số thực. Các giá trị này lập thành những khoảng liên tục trên trục số.

Ví dụ: Tập hợp chiều cao của giống lúa ST1 trong khoảng 100cm đến 110cm, chỉ tiêu chiều cao cũng là một biến ngẫu nhiên liên tục vì trong khoảng số này lấy ra một số bất kỳ đều có thể là chiều cao của một cây lúa nào đó của giống lúa ví dụ 105,2cm...

Tóm lại biến số ngẫu nhiên được mô tả trong Hình 1.1

Hình 1.2: Sơ đồ mô tả biến số ngẫu nhiên.



Thường các chỉ tiêu thu thập trong các thí nghiệm đa số là biến ngẫu nhiên liên tục như: hàm lượng đường trong nước mía, hàm lượng vitamin C trong trái cây, hàm lượng enzyme amylase trong hạt lúa, chiều cao của cây, năng suất cây trồng...

1.4. Các loại thang đo trong thống kê

Có 4 thang đo được dùng với các biến số: thang đo định danh, thang đo thứ bậc, thang đo khoảng và thang đo tỷ lệ. Thang đo định danh và thứ bậc gọi chung là thang đo định tính, thang đo khoảng và tỷ lệ gọi chung là thang đo định lượng.

1.4.1. Thang đo định danh

Là loại thang đo sử dụng cho các tiêu thức thuộc tính, không có sự hơn kém, không có thứ bậc. Người ta thường dùng các mã số để phân loại các đối tượng. Ngoài vai trò này, các mã số không mang ý nghĩa nào khác. Ví dụ:

- Giới tính: 1. Nam 2. Nữ
- Giống cây trồng: 1. Dưa hấu 2. Dưa lê 3. Dưa lưới 4. Loại dưa khác

- Màu sắc: 1. Xanh 2. Đỏ 3. Vàng

1.4.2. Thang đo thứ bậc

Là loại thang đo sử dụng các con số hoặc tự có sự hơn kém nhau, có thứ tự nhưng không có khoảng cách giữa các điểm khác nhau trong thang. Sự chênh lệch giữa các biểu hiện không nhất thiết phải bằng nhau trong thang.

Ví dụ:

Thẻ trạng của vật nuôi:

1. Rất gầy
2. Gầy,
3. Trung bình,
4. Béo
5. Rất béo.

Mức độ độc hại của chất amiăng đối với công nhân:

1. Thấp
2. Trung bình,
3. Cao

1.4.3. Thang đo khoảng

Thang đo khoảng thường dùng cho các đặc điểm số lượng và đôi khi cũng được áp dụng cho các đặc điểm thuộc tính. Thang đo khoảng là thang đo thứ bậc có các khoảng cách đều nhau. Các phép tính cộng trừ đều có ý nghĩa.

Ví dụ: Ông (Bà) cho biết ý kiến của mình về một số đặc điểm của giống lúa IR504 qua hai năm canh tác:

- Năng suất: 1: Rất thấp 2: thấp 3: Trung bình 4: cao 5: rất cao

1.4.4. Thang đo tỷ lệ

Là loại thang đo dùng cho đặc tính số lượng. Thang đo tỷ lệ có đầy đủ các đặc tính của thang đo khoảng.

2. Dữ liệu

2.1. Khái niệm

Để nghiên cứu một vấn đề nào đó ta phải quan sát, ghi nhận, thu thập,... các thuộc tính, số đo, số lượng, phản ánh bản chất của nó. Các quá trình đó sẽ tạo ra một tập dữ liệu cho vấn đề quan tâm. Dữ liệu thường được đo ở dạng thang số hoặc phân loại thành nhóm rồi sau đó mã hóa dưới dạng số. Vấn đề quan trọng của việc thu thập dữ liệu là xác định rõ ràng những dữ liệu nào cần thu thập, thứ tự ưu tiên của các loại dữ liệu này. Vấn đề nghiên cứu và mục tiêu nghiên cứu càng cụ thể thì việc xác định dữ liệu cần thu thập càng dễ dàng. Dữ liệu luôn đóng một vai trò vô cùng quan trọng trong nghiên cứu thống kê ứng dụng.