

UBND TỈNH SƠN LA
TRƯỜNG CAO ĐẲNG SƠN LA

GIÁO TRÌNH
MÔN HỌC/MÔ ĐUN: XỬ LÝ THỐNG KÊ TOÁN HỌC
TRONG LÂM NGHIỆP
NGÀNH/NGHỀ: LÂM SINH
TRÌNH ĐỘ: TRUNG CẤP.

2 Tín chỉ (Lý thuyết: 15 giờ; Thực hành: 43 giờ; Kiểm tra: 02 giờ)

Ban hành kèm theo Quyết định số: 628 /QĐ-CĐSL, ngày 08 /09/ 2023
của Hiệu trưởng Trường Cao Đẳng Sơn La về việc ban hành Chương trình
đào tạo ngành/ nghề Lâm sinh, trình độ cao đẳng và trung cấp.

Sơn La, năm 2023

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.

LỜI GIỚI THIỆU

Để đáp ứng yêu cầu giảng dạy chương trình đào tạo trình độ trung cấp nghề lâm sinh, tôi đã tiến hành biên soạn bài giảng môn Xử lý thống kê toán học trong lâm nghiệp. Thống kê toán học trong lâm nghiệp là môn học nghiên cứu những kiến thức cơ bản về lý thuyết và thực hành cách tính toán số liệu điều tra lâm nghiệp để học các môn chuyên môn của nghề

Bài giảng gồm 4 chương, cấu trúc thống nhất phù hợp với phương pháp giảng dạy tích hợp; Bài 1: Lập phân bố thực nghiệm và quan sát các đặc trưng mẫu, Bài 2: Mô hình hóa quy luật cấu trúc tần số, Bài 3: So sánh các mẫu thí nghiệm và quan sát trong lâm nghiệp, Bài 4: Phân tích tương quan hồi quy.

Để biên soạn bài giảng này tôi đã tham khảo nhiều tài liệu, lấy ý kiến của nhiều chuyên gia là các nhà nghiên cứu, các cơ sở sản xuất kinh doanh, giáo viên có kinh nghiệm giảng dạy trong và ngoài trường.

Vì thời gian có hạn, tuy đã có nhiều cố gắng nhưng không thể tránh khỏi thiếu sót, rất mong được bạn đọc góp ý để bài giảng hoàn thiện hơn.

Sơn La, tháng 09 năm 2023

Giảng viên biên soạn : Th s Lê Thị Hạnh.

MỤC LỤC

TUYÊN BỐ BẢN QUYỀN

LỜI GIỚI THIỆU

Bài 1: Lập phân bố thực nghiệm và quan sát.....	1
các đặc trưng mẫu	1
A. MỤC TIÊU	1
B. NỘI DUNG	1
1.1. Những khái niệm cơ bản	1
1.2. Các đặc trưng mẫu	3
C. CÂU HỎI VÀ BÀI TẬP THỰC HÀNH	8
1. Câu hỏi đánh giá kiến thức	8
2. Bài tập rèn luyện kỹ năng	8
D. GHI NHỚ	9
Bài 2: Mô hình hóa quy luật cấu trúc tần số.	10
A. MỤC TIÊU	10
B. NỘI DUNG	10
3.1. Một số khái niệm cơ bản về kiểm định giả thuyết thống kê.	10
3.2. ý nghĩa của việc kiểm định các quy luật cấu trúc tần số	11
3.4. Kiểm định giả thuyết một số phân bố lý thuyết thường gặp trong lâm nghiệp	12
C. CÂU HỎI VÀ BÀI TẬP THỰC HÀNH	20
1. Câu hỏi đánh giá kiến thức	20
2. Bài tập rèn luyện kỹ năng	20
D. GHI NHỚ	20
Bài 3: SO SÁNH CÁC MẪU THÍ NGHIỆM VÀ QUAN SÁT	21
TRONG LÂM NGHIỆP.	21
A. MỤC TIÊU	21
B. NỘI DUNG	21
4.1. ý nghĩa	21

4.2. Tr- ờng hợp các mẫu độc lập	22
4.3 Tr- ờng hợp các mẫu liên hệ về l- ợng	27
4.4. So sánh các mẫu độc lập về chất (các biến định tính) -	30
4.5 Kiểm tra quan hệ giữa 2 biến định tính dựa trên bảng chéo (Crosstabs)	32
C. CÂU HỎI VÀ BÀI TẬP THỰC HÀNH	34
1. Câu hỏi đánh giá kiến thức	34
2. Bài tập rèn luyện kỹ năng	34
D. GHI NHỚ	35
Bài 4: PHÂN TÍCH TƯƠNG QUAN HỒI QUY.	36
A. MỤC TIÊU	36
B. NỘI DUNG	36
6.1 Các mô hình hồi quy	36
6.2 Các đặc tr- ng chỉ mức độ liên hệ giữa các đại l- ợng	37
6.3. Hồi quy tuyến tính một lớp	46
C. CÂU HỎI VÀ BÀI TẬP THỰC HÀNH	49
1. Câu hỏi đánh giá kiến thức	49
2. Bài tập rèn luyện kỹ năng	50
D. GHI NHỚ	51
HƯỚNG DẪN GIẢNG DẠY BÀI GIẢNG	52
I. Vị trí, tính chất của môn học:	52
II. Mục tiêu môn học:	52
III. Nội dung môn học:	52
TÀI LIỆU THAM KHẢO	54

BÀI 1: LẬP PHÂN BỐ THỰC NGHIỆM VÀ QUAN SÁT CÁC ĐẶC TRƯNG MẪU

(Lý thuyết: 3 giờ; Thực hành: 12 giờ; Kiểm tra: 0 giờ)

A. MỤC TIÊU

* Kiến thức:

- Trình bày được khái niệm tổng thể, mẫu quan sát, các dấu hiệu quan sát, bố trí mẫu điều tra

Liệt kê được các bước tính toán thống kê cho đại lượng liên tục và đại lượng đứt quãng

Trình bày được các đặc trưng của phân bố thực nghiệm

* Kỹ năng:

- giải được các dạng bài tập phân bố thực nghiệm.

* Năng lực tự chủ và trách nhiệm:

Có khả năng làm việc độc lập khi tính toán và viết một báo cáo điều tra thống kê, kiểm kê rừng

Có khả năng chịu trách nhiệm về mức độ chính xác trong công việc điều tra tài nguyên rừng

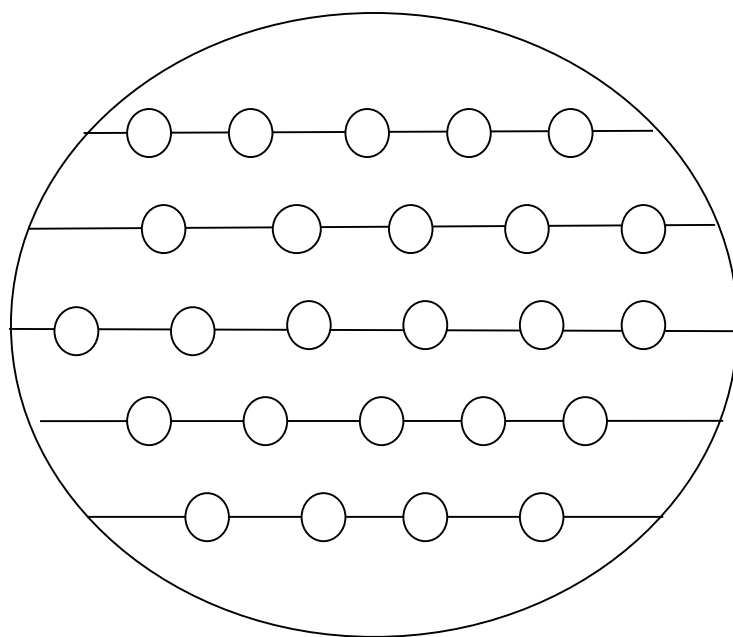
B. NỘI DUNG

1.1. Những khái niệm cơ bản

1.2 Khái niệm về tổng thể và mẫu

Tổng thể theo định nghĩa chung là một tập hợp hữu hạn hoặc vô hạn các phần tử có cùng một số tính chất chung nào đó . Chẳng hạn tập hợp tất cả các cây rừng trong một khu rừng rộng lớn . Tính chất chung ở đây là cây rừng chứ không phải là tre nứa hoặc các loại động vật .Ng-ời ta th-ờng ký hiệu N là số phần tử trong tổng thể . Cũng cần nói thêm rằng trong điều tra trừ 1- ợng $N =$ diện tích rừng / diện tích ô quan sát đặt theo hệ thống hoặc ngẫu nhiên . Còn mẫu là một bộ phận đ- ợc chọn từ tổng thể theo một ph- ơng pháp nào đó . Dung 1- ợng mẫu th- ờng ký hiệu n ($n < N$) . Th- ờng có 3 cách chọn mẫu đ- ợc dùng trong Lâm nghiệp

- Chọn ngẫu nhiên : Các phần tử tổng thể đ- ọc đánh số và dùng cách rút thăm hoặc bảng ngẫu nhiên để chọn ra n phần tử quan sát. Các phần tử có thể chọn một lần (không hoàn lại) hoặc có hoàn lại . Nếu $N > 10n$ thì việc chọn có hoàn lại và không hoàn lại là nh- nhau khi tính sai số rút mẫu .Ph- ơng pháp này có - u điểm là khách quan để thực hiện , nh- ng các phần tử ở mẫu có thể không phân bố đều trong tổng thể
- Chọn hệ thống : Đây là ph- ơng pháp th- ờng đ- ọc dùng trong Lâm nghiệp nhất là trong điều tra rừng .ở ph- ơng pháp này, trên diện tích rừng ng- ời ta kẻ nhiều đ- ồng song song cách đều và trên đó đặt những ô cách đều có diện tích nh- nhau để tiến hành quan sát các đại l- ợng nh- đ- ồng kính , chiềucao hoặc trữ l- ợng cây gỗ vv.(Xem hình1.1



Hình 1.1 Ô hệ thống cách đều theo tuyến

Ph- ơng pháp này có - u điểm là các phần t- ở mẫu rải đều trong tổng thể tính đại diệ của mẫu cao . Nh- ng có nh- ợc điểm là tính hệ thống sẽ bị vi phạm nếu gặp các ch- ớng ngại vật khi mở tuyến và đặt ô quan sát.

- Chọn mẫu điển hình : Trong một khu rừng ng- ời ta chọn hẳn cả một giải rừng mang tính chất điển hình cho đại l- ợng quan sát để thu thập số liệu . Ph- ơng pháp này đơn giản dễ thực hiện , nh- ng ít khách quan , độ chính xác phụ thuộc vào kinh nghiệm của điểuta viên . Ph- ơng pháp này **không tính đ- ợc sai số chọn mẫu**

Ngoài các phương pháp trên còn có các phương pháp chọn mẫu mang tính chất phức hợp như: mẫu phân khối (Stratified Sampling) mẫu phân cấp (Stage Sampling) và mẫu nhiều giai đoạn (multiphase Sampling) sẽ được trình bày kỹ hơn trong chương 10. Hiện nay Viện ĐTQH rừng đang áp dụng kiểu mẫu 2 cấp đặt hệ thống để tiến hành điều tra đánh giá và theo dõi diễn biến tài nguyên rừng toàn quốc theo chu kỳ 5 năm một lần

1.2 Các đặc trưng mẫu

1.2.1 – Khái niệm chung về số đặc trưng mẫu

Bảng và biểu đồ cho ta biết một cách khái quát về quy luật biến thiên của dấu hiệu quan sát. Nhiều khi chúng ta cần biết những số rất điển hình cho từng mặt của quy luật biến thiên ấy. Những số này gọi chung là những số đặc trưng mẫu. Tùy theo tính chất đặc trưng của nó người ta chia làm 3 loại: các số đặc trưng vị trí, các số đặc trưng về biến động và các đặc trưng về hình dạng. Sau đây chúng ta sẽ lần lượt xét những đặc trưng quan trọng nhất của ba loại đặc trưng trên.

1.2.2 – Những đặc trưng vị trí

1.2.2.1 – Số bình quân cộng của mẫu (gọi tắt là số bình quân hay số trung bình mẫu):

1.2.2.1.1. Định nghĩa: Giả sử có một dãy trị số quan sát $x_1, x_2, x_3, \dots, x_n$ thì trị số

$$\bar{x} = \frac{1}{n} \cdot (x_1 + x_2 + x_3 + \dots + x_n)$$

$$\text{hay } \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (1-1)$$

gọi là số bình quân cộng giản đơn. Số bình quân này thường tính với tài liệu quan sát có dung lượng mẫu nhỏ ($n < 30$) chưa qua chỉnh lý.

Ví dụ: Dãy trị số quan sát 10,3, 10,7, 12,4, 11,5, 12,6, 14,1, 12,4, 14,5, 12,2, 13,8.

$$\bar{x} = \frac{1}{10} \cdot (10,3 + 10,7 + 12,4 + 11,5 + 12,6 + 14,1 + 12,4 + 14,5 + 12,2 + 13,8) = \frac{124,5}{10} = 12,45$$

Nếu tài liệu đã qua chỉnh lý với $x_1, x_2, x_3, \dots, x_m$ là các trị số giữa tổ có tần số tương ứng là $f_1, f_2, f_3, \dots, f_m$ thì số bình quân cộng \bar{x} được tính theo công thức:

$$\bar{x} = \frac{1}{n} \cdot (f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_m \cdot x_m)$$

$$\text{hay } \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^m f_i \cdot x_i \quad (1.2)$$

gọi là *số bình quân gia quyền*.

Ví dụ: số bình quân gia quyền tính theo bảng (1.5)

$$\bar{x} = \frac{1}{50} \cdot (1 \cdot 6,50 + 2 \cdot 7,0 + 5 \cdot 7,5 + \dots + 1 \cdot 10) = 8,37m$$

1.2.2.1. 2. Một vài tính chất của số bình quân cộng:

- Tổng chênh lệch từ các trị số quan sát đến số bình quân cộng bằng không:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Chứng minh:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n \cdot \bar{x} \text{ nh-ng theo (1-1) thì } \sum_{i=1}^n x_i = n \cdot \bar{x}$$

$$\text{do đó } \sum_{i=1}^n (x_i - \bar{x}) = n \bar{x} - n \bar{x} = 0$$

Ví dụ theo dãy (1)

$$(10 - 12,45) + (10,7 - 12,45) + \dots + (13,8 - 12,45) = 0$$

- Tổng các chênh lệch bình phương từ các trị số đến số bình quân nhỏ hơn tổng các chênh lệch bình phương từ các trị số đến bất kỳ một số A nào đó (với $A \neq \bar{X}$).

$$\sum_{i=1}^n (x_i - \bar{x})^2 < \sum_{i=1}^n (x_i - A)^2$$

Chứng minh:

$$\begin{aligned} \sum_{i=1}^n (x_i - A)^2 &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - A)]^2 \\ &= \sum_{i=1}^n \left[(x_i - \bar{x})^2 + 2 \cdot (x_i - \bar{x})(\bar{x} - A) + (\bar{x} - A)^2 \right] \\ \sum_{i=1}^n (x_i - A)^2 &= \sum_{i=1}^n \left[(x_i - \bar{x})^2 + 2 \cdot (\bar{x} - A) \sum_{i=1}^n (x_i - \bar{x}) + \sum_{i=1}^n (\bar{x} - A)^2 \right] \end{aligned}$$

$$(1) \qquad (2) \qquad (3)$$

Theo tính chất thứ nhất của số bình quân cộng ở trên thì số hạng (2) trong biểu thức trên sẽ triệt tiêu, vì vậy:

$$\sum_{i=1}^n (x_i - A)^2 = \sum_{i=1}^n \left[(x_i - \bar{x})^2 + n(\bar{x} - A)^2 \right]$$

Tức là:
$$\sum_{i=1}^n (x_i - A)^2 > \sum_{i=1}^n (x_i - \bar{x})^2$$

- Nếu đem các giá trị của đại lượng X cộng hoặc trừ cho một hằng số thì số bình quân của đại lượng mới cũng sẽ bằng bình quân của đại lượng X cộng hay trừ với hằng số đó.

Chứng minh: Đặt $Z = X \pm A$

Theo định nghĩa về số bình quân ta có:

$$\bar{z} = \frac{1}{n} \cdot \sum_{i=1}^n z_i \text{ thay } Z = X \pm A \text{ ta có:}$$

$$\bar{z} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i \pm A) = \frac{1}{n} \cdot \sum_{i=1}^n x_i \pm A$$

$$\bar{z} = \bar{x} \pm A$$

Số bình quân được vận dụng rất rộng rãi trong lâm nghiệp, nhất là trong công tác điều tra đo đếm trữ sản lượng rừng.

Chẳng hạn để xác định thể tích cây đứng ng-ời ta thường dùng công thức $V = g.h.f$. Trong đó g là tiết diện ngang của thân cây ở độ cao 1,3m, h là chiều cao của cây có thể đo bằng dụng cụ đo cao, f là tỷ số giữa thể tích thực của cây và thể tích viên trụ có cùng tiết diện ngang ở độ cao 1,3m và chiều cao bằng chiều cao cây, tỷ số này (f) gọi là *hình số ngang ngược* và rất khó xác định ở những cây đứng riêng lẻ. Ng-ời ta thường ở một loài cây xác định thì quy luật phân bố của hình số thường có dạng đối xứng và có thể thay hình số của các cây riêng lẻ bằng hình số trung bình đã xác định được trên cơ sở những nghiên cứu trước đây, hoặc ở những khu vực có điều kiện tương tự. Như vậy khi tính thể tích của toàn bộ lâm phần ($\sum V = \sum g.h.f$) thì kết quả vẫn không có sai số lớn so với kết quả tính toán hình số riêng cho từng cây. Ng-ời ta đã lợi dụng tính chất này để lập các biểu thể tích 2 nhân tố. Có nghĩa là ở biểu thể tích này chỉ cần quan sát chiều cao và đường kính, còn hình số $f_{1.3}$ thì lấy một giá trị trung bình nào đó.

1.5.2.2– Số trung bình toàn phương