

Vận dụng phần mềm Statgraphics trong giảng dạy thống kê cho sinh viên ngành Nông nghiệp

Trần Thị Ngọc Giàu*

* Trường Đại học An Giang - Đại học Quốc Gia TP.HCM

Received: 9/2/2023; Accepted: 13/2/2023; Published: 9/3/2023

Abstract: The application of informatics in statistical data processing is increasingly being applied more and more widely in various fields, including the field of Agriculture. There are many specialized softwares for statistical processing, depending on the field, you can choose the suitable software. The content of the article is the first step to create conditions for students majoring in Agriculture to access and apply Statgraphics software in teaching Probability – statistics subject. In addition, students can use this specialized software in solving specialized practical problems.

Keywords: Statgraphics software, statistical data, data analysis.

1. Đặt vấn đề

Xử lý thống kê thông qua công nghệ tin học ngày nay đã phát triển rất lâu và đóng vai trò quan trọng, giúp cho con người rút ngắn được thời gian tính toán và xử lý được một lượng lớn thông tin. Có rất nhiều phần mềm thống kê chuyên dụng và phổ biến như Excel, SPSS, Statgraphics, ... bao gồm các xử lý thống kê phổ biến như: phân tích các đặc trưng mẫu, so sánh các trung bình, tỉ lệ, phân tích phương sai, hồi quy một chiều, nhiều chiều, ... Riêng đối với lĩnh vực Nông nghiệp, phần mềm Statgraphics được sử dụng phổ biến nhất bởi một số tính năng nổi trội trong phân tích dữ liệu, thuận tiện trong hoạt động quản lý và nghiên cứu lâm nghiệp, quản lý tài nguyên thiên nhiên, môi trường. Chính vì thế, trong quá trình giảng dạy môn Xác suất thống kê cho sinh viên ngành Nông nghiệp, giảng viên cần tạo điều kiện cho sinh viên tiếp cận với phần mềm chuyên dụng này thông qua các bài thực hành tính toán số liệu, giúp sinh viên làm quen với phần mềm trong xử lý số liệu. Bên cạnh đó, việc sử dụng thành thạo phần mềm chuyên dụng là điều kiện để sinh viên có thể học tốt học phần Phương pháp thống kê chuyên ngành sau này.

2. Nội dung nghiên cứu

2.1. Tổng quan phần mềm Statgraphics Centurion phân tích thống kê

Đây là một phần mềm chuyên dụng trong xử lý thống kê, bao gồm các chức năng sau:

- Tạo lập cơ sở dữ liệu dưới dạng bảng tính;
- Tính toán các đặc trưng mẫu, vẽ sơ đồ, đồ thị quan hệ.
- So sánh hai hay nhiều mẫu bằng các tiêu chuẩn

thống kê và nhiều tiêu chuẩn phi tham số khác.

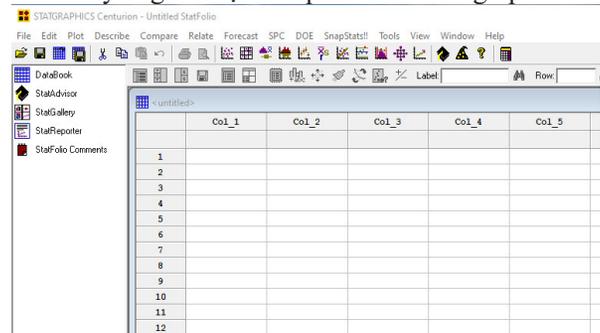
- Phân tích phương sai ANOVA.

- Thiết lập các mô hình hồi quy tuyến tính hay phi tuyến tính từ một cho đến nhiều lớp, tổ hợp biến.

So với Excel, phần mềm Statgraphics có ưu điểm hơn: dữ liệu được trực quan hóa, khả năng phân tích đồ họa cao, tích hợp với giao diện R và giúp chia sẻ đầy nhanh nguồn thông tin. Có hơn 150 phép phân tích trên Menu chính của phần mềm Statgraphics. Tuy nhiên, trong giới hạn bài viết này, tác giả bước đầu giúp sinh viên ngành Nông nghiệp làm quen với phần mềm này thông qua một số bài toán thống kê cơ bản và đã vận dụng trong quá trình giảng dạy học phần Xác suất thống kê. Đó cũng là kiến thức nền tảng để sinh viên ngành Nông nghiệp thành thạo trong việc xử lý số liệu trong quá trình lĩnh hội kiến thức chuyên ngành sau này.

Việc cài đặt phần mềm Statgraphics khá đơn giản.

Dưới đây là giao diện của phần mềm Statgraphics:



Đối với dữ liệu, chúng ta có thể nhập trực tiếp dữ liệu hoặc mở dữ liệu từ phần mềm Excel. Tuy nhiên, dữ liệu lập trong Excel cần lưu dưới dạng phiên bản của Excel 97 – 2003 với đuôi là .xls.

Đối với phần mềm Statgraphics Centurion, kết quả phân tích sẽ được hiển thị gồm 4 cửa sổ: 2 cửa sổ bên trái là output dạng bảng và 2 cửa sổ bên phải là output dạng đồ thị. Đối với phần mềm Statgraphics Plus, kết quả được hiển thị chỉ gồm 2 cửa sổ: bên trái là output dạng bảng và bên phải là output dạng đồ thị. Nếu double-click đôi vào một cửa sổ bất kỳ thì sẽ được phóng lớn ở kích thước cực đại để dễ dàng quan sát và đọc kết quả, double-click đôi tiếp tục vào cửa sổ đó để trở về màn hình ban đầu.

2.2. Thực hành xử lý số liệu thống kê bằng phần mềm Statgraphics

a. Phân tích các đặc trưng mẫu – Kiểm định giả thuyết:

Bài toán 1: Thống kê số quả chấu/cây của 2 giống cà chua A, B được trình bày ở bảng 2.1:

Bảng 2.1

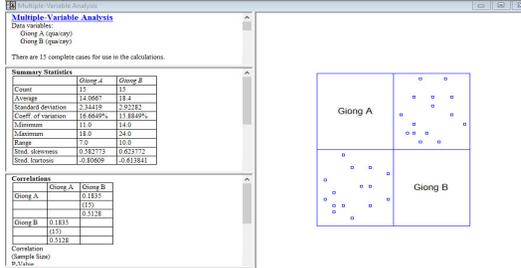
Giống A	12	14	11	15	14	13	12	16	18	16	18	16	13	12	11
Giống B	15	22	20	19	14	16	18	19	21	22	17	16	24	16	17

Xác định các giá trị trung bình, phương sai, độ lệch chuẩn, sai số trung bình và hệ số biến động của tính trạng số quả/cây giữa 2 giống cà chua trên.

Cách thực hiện:

- Nhập dữ liệu;
- Vào menu Describe/Numeric Data/Multiple Valuable Analysis ...
- Xuất hiện hộp thoại, chọn cả 2 giống A và B cho Data, nhấn OK.

Kết quả như sau:

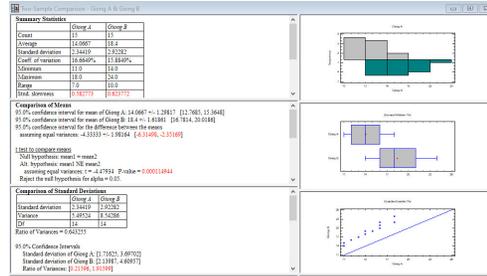


Kết quả thể hiện trên 2 cửa sổ. Cửa sổ bên trái cho biết thông tin số liệu bảng thống kê của 2 mẫu dữ liệu, cửa sổ bên phải là biểu đồ. Nếu muốn mở to cửa sổ nào, ta chỉ cần double-click vào cửa sổ ấy. Chúng ta quan tâm đến các số liệu ở cửa sổ bên trái, thể hiện kết quả các giá trị thống kê của 2 mẫu dữ liệu. Nếu muốn hiển thị thêm các giá trị khác thì click chuột phải vào cửa sổ đó, chọn Pane Option, sau đó click chọn giá trị muốn hiển thị và nhấn OK.

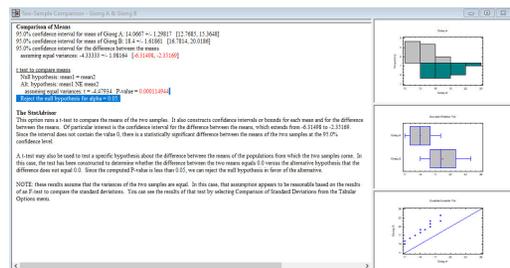
Bài toán 2: Với dữ liệu ở bài toán 1. Hãy so sánh số quả chấu/cây của 2 giống cà chua A và B với mức

ý nghĩa 5%. Giải thích kết quả.

Cách thực hiện: Từ màn hình Statgraphics với dữ liệu đã nhập, vào Compare/ Two Samples/ Independent Sample... nhấn OK. Kết quả thể hiện như sau:



Ở đây, chúng ta quan tâm đến các số liệu cửa sổ bên trái. Bảng trên cho kết quả các đặc trưng mẫu, tương tự bài toán 1. Trong phần đầu cho kết quả các khoảng tin cậy 95% của trung bình của mỗi giống, khoảng tin cậy 95% về sự chênh lệch giữa hai trung bình của hai giống cà chua. Với bài toán so sánh hai trung bình như yêu cầu, chúng ta sử dụng t-test, giá trị của thống kê là $t = -4,47934$ và kết luận được thể hiện ở dòng cuối: **Reject the null hypothesis for alpha = 0.05**, nghĩa là bác bỏ giả thuyết H_0 ở mức ý nghĩa 5%. Nếu chúng ta muốn quan tâm đến kết quả so sánh độ lệch tiêu chuẩn của hai mẫu trên thì có thể theo dõi số liệu sau.



c. Phân tích hồi quy

Bài toán 3: Người ta dùng ba mức nhiệt độ gồm 105°C, 120°C, 135°C kết hợp với ba khoảng thời gian là 15, 30 và 60 phút để thực hiện một phản ứng tổng hợp. Các hiệu suất của phản ứng (%) được trình bày trong bảng 2.2:

Bảng 2.2

Thời gian X_1 (phút)	Nhiệt độ X_2 (°C)	Hiệu suất Y (%)
15	105	1.87
30	105	2.02
60	105	3.28
15	120	3.05
30	120	4.07
60	120	5.54
15	135	5.03

30	135	6.45
60	135	7.26

Hãy cho biết yếu tố nhiệt độ và/hoặc thời gian có liên quan tuyến tính với hiệu suất của phản ứng tổng hợp hay không?

Cách thực hiện:

Bước 1. Nhập số liệu:

	Thời gian	Nhiệt độ	Hiệu suất
	X1	X2	Y
1	15	105	1.87
2	30	105	2.02
3	60	105	3.28
4	15	120	3.05
5	30	120	4.07
6	60	120	5.54
7	15	135	5.03
8	30	135	6.45
9	60	135	7.26
10			

Bước 2. Chọn Relate/One Factor/ Simple Regression để ấn định các giá trị của biến số Y và X.

Bước 3. Nhấn OK, kết quả phân tích sẽ được hiển thị trong 4 cửa sổ.

Parameter	Estimate	Error	Statistic	P-Value
Intercept	-11.1407	1.28011	-2.12903	0.0708
Slope	0.0445597	0.0322708	1.38019	0.2100

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-Value
Model	6.24892	1	6.24892	1.90	0.2100
Residual	22.9629	7	3.28041		
Total (Corrected)	29.2118	8			

Row	X	Y	Studentized Residual	Residual

Chúng ta quan tâm đến số liệu tô đỏ trong bảng kết quả ở cửa sổ bên trái. Dựa vào bảng kết quả phân tích, chúng ta có nhận xét sau:

- Phương trình hồi quy hiệu suất - thời gian: $Y = 2.73 + 0.044X_1$.
- $P\text{-value} = 0.0708 > 0.05$: Suy ra, hệ số a không có nghĩa (với độ tin cậy 95%).
- $P\text{-value} = 0.21 > 0.05$: Suy ra, hệ số b không có nghĩa (với độ tin cậy 95%).

Vậy, phương trình hồi quy không có ý nghĩa thống kê. Do đó, có thể kết luận yếu tố thời gian không có liên quan tuyến tính với hiệu suất phản ứng.

Tương tự, có thể tìm được phương trình hồi quy – nhiệt độ và dự đoán được mối tương quan giữa hai

biến này, kết quả như sau:

Parameter	Estimate	Error	Statistic	P-Value
Intercept	-11.1411	1.23966	-4.43189	0.0012
Slope	0.128156	0.0270254	4.73719	0.0021

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-Value
Model	22.3108	1	22.3108	22.63	0.0021
Residual	6.90101	7	0.985858		
Total (Corrected)	29.2118	8			

- Phương trình hồi quy hiệu suất – nhiệt độ: $Y = -11.14 + 0.128X_2$.
- $P\text{-value} = 0.011 < 0.05$: Hệ số a có nghĩa (với độ tin cậy 95%).
- $P\text{-value} = 0.002 < 0.05$: Hệ số b có nghĩa (với độ tin cậy 95%).

Vậy phương trình hồi quy có ý nghĩa thống kê, nghĩa là yếu tố nhiệt độ có liên quan tuyến tính với hiệu suất phản ứng.

Đối với môn Xác suất thống kê, yêu cầu đạt được là sinh viên có thể tính toán và xử lý số liệu với những bài toán khá đơn giản với những dữ liệu cho sẵn. Tuy nhiên, đối với những bài toán chuyên ngành, ngoài số liệu nhận được ở cửa sổ bên trái thì kết quả từ đồ thị cần được quan tâm để đưa ra kết luận cho bài toán thống kê.

3. Kết luận

Có rất nhiều phần mềm Tin học hỗ trợ rất mạnh cho việc xử lý số liệu nghiên cứu, mỗi phần mềm có đặc trưng, điểm mạnh, điểm yếu riêng. Tuy nhiên, đối với lĩnh vực Nông nghiệp, Statgraphics có thể nói là phần mềm mạnh nhất để phân tích dữ liệu, trực quan hóa dữ liệu, mô hình thống kê và dự báo. Việc giảng dạy môn Xác suất thống kê có lồng ghép một số bài thực hành trên Statgraphics giúp cho sinh viên có nhiều hứng thú hơn trong việc tiếp thu kiến thức mới, sinh viên có cơ hội làm quen với phần mềm chuyên dụng và tiến tới giải quyết được những vấn đề thực tiễn chuyên ngành sau này.

Tài liệu tham khảo

1. Lê Sĩ Đồng (2010). *Xác suất thống kê và ứng dụng*. NXB Giáo dục Việt Nam.
1. Bảo Huy (2017). *Tin học thống kê trong Nông nghiệp*. NXB Khoa học Kỹ thuật.
2. Phan Thanh Kiểm (2010). *Cơ sở của Toán học của phép xử lý thống kê trong nghiên cứu khoa học Nông nghiệp*. NXB Nông Nghiệp.
3. Lê Đức Vĩnh (2006). *Giáo trình Xác suất thống kê*. Đại học Nông Nghiệp 1- Hà Nội.