

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



GIÁO TRÌNH

**MÔN HỌC: XÁC XUẤT THỐNG KÊ
NGÀNH, NGHỀ: DỊCH VỤ THÚ Y
TRÌNH ĐỘ: CAO ĐẲNG**

*(Ban hành kèm theo Quyết định Số: .../QĐ-CĐCĐ-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: Thống kê thường được chia thành hai lãnh vực:
- *Thống kê mô tả (Descriptive Statistics): bao gồm các phương pháp thu thập số liệu, mô tả và trình bày số liệu, tính toán các đặc trưng đo lường.* - *Thống kê suy diễn (Inferential Statistics): bao gồm các phương pháp như ước lượng, kiểm định, phân tích mối liên hệ, dự đoán. trên cơ sở các thông tin thu thập từ mẫu.*

Một số thuật ngữ dùng trong bố trí thí nghiệm n đơn vị thí nghiệm (experimental unit): vật liệu à tác động một hoặc một số nhân tố là đo lường các ảnh hưởng của nó. nhân tố (factor) là nguyên nhân gây ảnh hưởng đến các giá trị quan sát là bao gồm các mức độ khác nhau. nghiệm thức (treatment) có thể bao gồm các mức độ khác nhau của một nhân tố hoặc một phối hợp các mức độ của các nhân tố khác nhau mà ta muốn khảo sát ảnh hưởng của nó trên vật liệu thí nghiệm. sai số thí nghiệm (experimental error) là tổng cộng các nguồn biến động không kiểm soát được. Nguồn biến động luôn hiện hữu trong vật liệu thí nghiệm do phương pháp thực hiện thí nghiệm hoặc do người làm thí nghiệm.

Mặc dù đã rất cố gắng trong quá trình biên soạn, song không tránh khỏi những sai sót. Chúng tôi rất mong nhận được những đóng góp ý kiến của các bạn và đồng nghiệp để cuốn giáo trình hoàn thiện hơn.

Đồng Tháp, ngày.....tháng... năm 2017

Chủ biên: Ths.Cao Thanh Hoàn

MỤC LỤC

LỜI GIỚI THIỆU	ii
CHƯƠNG 1.....	1
NHỮNG VẤN ĐỀ CƠ BẢN CỦA THỐNG KÊ SINH HỌC	1
1. Một số khái niệm cơ bản	1
1.1. Tổng thể ($n \geq 30$).....	1
1.2. Mẫu ($n < 30$).....	2
1.3. Tính trạng	2
1.4. Số liệu.....	2
1.5. Biến số, biến thiên và tham số	3
2. Các dạng biểu đồ thường gặp.....	3
2.1. Biểu đồ hình quạt	3
2.2. Biểu đồ hình cột	4
2.3. Biểu đồ hình gấp khúc.....	6
3. Các tham số đặc trưng của mẫu	9
3.1. Số trung bình cộng	9
3.2. Phương sai.....	9
3.3. Độ lệch chuẩn.....	9
3.4. Hệ số biến thiên.....	9
3.5. Sai số trung bình.....	10
4. Phép thử T	10
4.1. Kiểm tra mức độ tin cậy của số trung bình mẫu so với số trung bình của tổng thể.....	10
4.2. So sánh hai số trung bình mẫu với nhau	11
4.3. So sánh tỉ lệ trung bình quan sát với tỉ lệ trung bình lý thuyết	11
5. Phép thử F	13
5.1. Trường hợp hai số trung bình mẫu tương đương nhau	13
6. Thực hành.....	19
5.2. Trường hợp so sánh nhiều hơn hai số trung bình. (tham khảo thêm).....	19
CHƯƠNG 2.....	22
NGUYÊN TẮC BỐ TRÍ THÍ NGHIỆM	22
1. Các khái niệm cơ bản	22
1.1. Bố trí thí nghiệm	22
1.2. Yếu tố thí nghiệm.....	22
1.3. Mức độ	22

1.4. Nghiệm thức	22
1.5. Đơn vị thí nghiệm	23
1.6. Sự lặp lại.....	23
1.7. Ngẫu nhiên hóa.....	23
1.8. Chia khối	23
1.9. Xác định vấn đề nghiên cứu.....	24
2. Nguyên tắc bố trí một thí nghiệm	24
2.1. Xác định mục tiêu nghiên cứu	25
2.2. Xác định các yếu tố thí nghiệm.....	25
2.3. Xác định các lô thí nghiệm (mức độ).....	25
2.4. Xác định các đơn vị thí nghiệm	26
2.5. Xác định sự quan sát	26
2.6. Xác định mẫu thí nghiệm	26
2.7. Việc thực hiện thí nghiệm.....	27
2.8. Phân tích số liệu và giải thích kết quả.....	27
2.9. Viết báo cáo.....	27
3. Thực hành.....	27
CHƯƠNG 3.....	31
THÍ NGHIỆM MỘT YẾU TỐ.....	31
1. Mẫu hoàn toàn ngẫu nhiên	31
1.1. Mục đích thí nghiệm	31
1.2. Yếu tố thí nghiệm.....	31
1.3. Đơn vị thí nghiệm	32
1.4. Chỉ tiêu quan sát.....	32
1.5. Mẫu thí nghiệm	32
1.6. Phân tích.....	32
2. Mẫu khối hoàn toàn ngẫu nhiên	42
2.1. Mục đích thí nghiệm	42
2.2. Yếu tố thí nghiệm.....	42
2.3. Đơn vị thí nghiệm	42
2.4. Chỉ tiêu quan sát.....	43
2.5. Mẫu thí nghiệm	43
2.6. Phân tích.....	43
CHƯƠNG 4.....	51

THÍ NGHIỆM NHIỀU YẾU TỐ	51
1. Mẫu hoàn toàn ngẫu nhiên	51
1.1. Mục đích thí nghiệm	51
1.2. Yếu tố thí nghiệm.....	52
1.3. Đơn vị thí nghiệm	52
1.6. Phân tích.....	54
1.7. Kết luận	56
2. Mẫu khối hoàn toàn ngẫu nhiên	57
2.1. Mục đích thí nghiệm	57
2.2. Yếu tố thí nghiệm.....	57
2.3. Đơn vị thí nghiệm	57
2.4. Chỉ tiêu quan sát.....	58
2.5. Mẫu thí nghiệm	58
2.6. Phân tích.....	58
2.7. Kết luận	61
3. Thực hành.....	66
TÀI LIỆU THAM KHẢO.....	67

GIÁO TRÌNH MÔ HỌC

Tên môn học: XÁC XUẤT THỐNG KÊ

Mã môn học: CNN270

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

- Vị trí: là môn học chuyên môn tự chọn cho người học ngành Cao đẳng Dịch vụ thú y. Là môn học chuyên ngành được học sau môn thống kê sinh học, tin học và các môn chuyên môn khác.
- Tính chất: Là môn học nghiên cứu những vấn đề cơ bản nhất về nguyên tắc bố trí một thí nghiệm trong nghiên cứu, các kiểu bố trí thí nghiệm và thuật toán cho mỗi kiểu cũng như cách thức thu thập số liệu, phân tích kết quả thí nghiệm và viết báo cáo khoa học.
- Ý nghĩa và vai trò của môn học: Giáo trình này rất có ý nghĩa trong giảng dạy và học tập, góp phần quan trọng trong chương trình môn học của ngành. Giúp sinh viên sử dụng các phép tính ứng dụng trong so sánh các nghiệm thức về thu hoạch, giống vật nuôi, thành phần dinh dưỡng,.....

Mục tiêu của môn học/mô đun:

- Về kiến thức: hiểu biết cách bố trí thí nghiệm và ý nghĩa của kết quả thí nghiệm.
- Về kỹ năng:
 - + Bố trí được thí nghiệm trong việc nghiên cứu một vấn đề nào đó; Biết cách thu thập số liệu, phân tích và viết báo cáo kết quả đạt được.
 - + Thực hiện đúng một số phương pháp cơ bản nhất của thống kê sinh học trong phương pháp thí nghiệm.
 - + Sử dụng phần mềm xử lý thống kê excel và phần mềm xử lý thống kê Minitab để bố trí thí nghiệm và trình bày báo cáo khoa học.
 - + Sử dụng phần mềm xử lý thống kê excel để phân tích số liệu và giải thích kết quả thí nghiệm hai yếu tố; Sử dụng phần mềm xử lý thống kê Minitab để phân tích số liệu và giải thích kết quả thí nghiệm hai yếu tố.
- Về năng lực tự chủ và trách nhiệm: Nghiêm túc, trung thực, an toàn, chính xác trong việc thu thập số liệu.

Nội dung của môn học:

Số TT	Tên chương, mục	Thời gian (giờ)
-------	-----------------	-----------------

		Tổng số	Lý thuyết	Thực hành, thí nghiệm, bài tập	Kiểm tra (định kỳ)/Ôn thi, Thi kết thúc môn học
1	Chương 1: Những vấn đề cơ bản của thống kê sinh học	12	4	8	
2	Chương 2: Những nguyên tắc bố trí thí nghiệm	7	3	4	
3	Chương 3: Thí nghiệm một yếu tố	12	4	8	
4	Chương 4: Thí nghiệm nhiều yếu tố	11	3	8	
	* Ôn thi	1			1
	* Thi kết thúc môn học	1			1
	Cộng	45	14	28	3

CHƯƠNG 1

NHỮNG VẤN ĐỀ CƠ BẢN CỦA THỐNG KÊ SINH HỌC

MH18- 01

Giới thiệu:

Giới thiệu một vài vấn đề, khái niệm cơ bản để hiểu rõ hơn về thống kê, thí nghiệm. Thực hiện được các phép thử trong thí nghiệm trong chăn nuôi, thú y.

Mục tiêu:

- Kiến thức: Hiểu được một số vấn đề cơ bản nhất của thống kê sinh học làm cơ sở cho môn học phương pháp thí nghiệm.
- Kỹ năng: Thực hiện đúng một số phương pháp cơ bản nhất của thống kê sinh học trong phương pháp thí nghiệm.
- Năng lực tự chủ và trách nhiệm: Học tập nghiêm túc, sáng tạo; áp dụng kiến thức, kỹ năng đã học để ứng dụng cơ bản thống kê sinh học trong phương pháp thí nghiệm một cách hiệu quả cao, tránh sai sót.

1. Một số khái niệm cơ bản

1.1. Tổng thể ($n \geq 30$)

Tổng thể là tập hợp tất cả các đối tượng như người, vật, sự vật...có chung một số tính chất nhất định nào đó mà nhà nghiên cứu cần khảo sát. Tổng thể là có thực và có thể liệt kê ra. Ví dụ số heo nái trong các trại chăn nuôi công nghiệp ở phía Nam. Nhưng có cũng những số liệu chỉ giả thiết và không thể liệt kê được như số heo nái hiện có ở Việt Nam.

Đặc trưng của tổng thể là rất lớn, thậm chí là vô cùng lớn. Tổng thể được miêu tả bằng các tham số của tổng thể.

- Trung bình tổng thể $= \mu$

- Phương sai tổng thể $= \sigma^2$

Tổng thể thường vô hạn nên ta phải chọn một số phần tử của tổng thể để nghiên cứu rồi từ các giá trị đặc trưng của quần thể này ta suy đoán về các thông số của tổng thể.

Công việc này là nghiên cứu trên một số mẫu đại diện cho tổng thể.

Một tổng thể có N phần tử:

$$N = \{x_1, x_2, x_3, \dots, x_n\}$$

Với N : số lượng phần tử của tổng thể hay kích thước của tổng thể.

x : là giá trị của những phần tử mà ta khảo sát.

1.2. Mẫu ($n < 30$)

Chúng ta có thể chọn một mẫu (có dung lượng mẫu là n) từ tổng thể một cách ngẫu nhiên. Mẫu có $n = 20$ từ tổng thể có $N = 1000$. Mẫu n được chọn làm đại diện cho một tổng thể, nhưng cách chọn mẫu này không có gì đảm bảo là đã chọn được một mẫu đại diện. Vì vậy muốn có độ tin cậy cao cần phải có sự lặp lại nhiều lần trong việc chọn mẫu đại diện.

Nghiên cứu mẫu đại diện sẽ dễ dàng hơn, nhanh chóng hơn và ít tốn kém hơn nghiên cứu cả một tổng thể. ($n \ll N$), các tham số của mẫu:

- Trung bình mẫu = $\bar{x}$ $\rightarrow$ Trung bình tổng thể = μ

- Phương sai mẫu = s^2 $\rightarrow$ Phương sai tổng thể = σ^2

Từ các số đo của mẫu ta có thể sử dụng các giá trị đó để ước tính cho quần thể

- Trung bình mẫu = $\bar{x}$

- Phương sai mẫu = s^2

$$n = \{x_1, x_2, x_3, \dots, x_n\}$$

với n : số lượng phần tử của mẫu khảo sát.

x : là giá trị của những phần tử mà ta khảo sát.

Mục đích của những nghiên cứu là muốn biết các thông tin của tổng thể. Tuy nhiên do tổng thể quá lớn, chúng ta không thể quan sát hết các phần tử của tổng thể được nên các nhà nghiên cứu mới tiến hành khảo sát một số phần tử đại diện cho tổng thể. Công việc này gọi là 'lấy mẫu'.

Điều kiện để mẫu có thể đại diện cho tổng thể: Các phần tử của mẫu phải được chọn ngẫu nhiên và việc chọn mẫu phải đảm bảo tính độc lập.

1.3. Tính trạng

Tính trạng kiểu hình, hay tính trạng (*Trait, character*) là một biến thể đặc trưng về kiểu hình của một sinh vật có thể do di truyền, do môi trường hoặc là sự kết hợp của cả hai yếu tố trên.

Ví dụ: màu mắt là một *đặc trưng*, màu mắt xanh, nâu hay hạt dẻ là các tính trạng.

1.4. Số liệu

Việc thu thập đủ và đúng dữ liệu rất quan trọng. Bạn không thể có một nghiên cứu tốt nếu không có các dữ liệu tốt. Số liệu thu thập được phân ra làm nhiều loại như số liệu định tính, số liệu định lượng, số liệu tần số...

Số liệu định tính (số liệu thuộc tính): số liệu đối với thông tin không phải bằng số như giống, màu mắt, màu hoa, loại máu...

Số liệu đo lường (số liệu định lượng): là số liệu có được thông qua việc đo lường hay tính toán như chiều cao, trọng lượng, năng suất...

Số liệu tần số: là số liệu dựa trên số lượng cá thể quan sát trong các nhóm khác nhau,...

1.5. Biến số, biến thiên và tham số

Trong thống kê, các đối tượng nghiên cứu được gọi là các đơn vị quan sát. Trên đơn vị này, các đặc tính được quan sát hoặc đo đạc được gọi là các biến số. Trong mỗi đối tượng nghiên cứu, các giá trị số gán cho biến số được gọi là các quan sát hay các biến.

Thí dụ: để nghiên cứu huyết áp của các sinh viên trong một trường đại học, các nhà nghiên cứu đo huyết áp tối đa và tối thiểu cho từng sinh viên. Huyết áp tối đa và tối thiểu là các biến số, số đo huyết áp là các quan sát, các sinh viên là các đơn vị quan sát.

2. Các dạng biểu đồ thường gặp

Khi dữ liệu thu được dưới dạng thứ hạng hoặc thuộc tính, mỗi một quan sát sẽ trở thành các nhóm hoặc thứ hạng. Chúng ta dùng biểu đồ dạng cột hoặc dạng bánh để biểu diễn số hoặc phần trăm của từng nhóm.

Phương pháp đồ thị thống kê là phương pháp trình bày và phân tích các thông tin thống kê bằng các biểu đồ, đồ thị và bản đồ thống kê. Phương pháp đồ thị thống kê sử dụng con số kết hợp với hình vẽ, đường nét và màu sắc để trình bày các đặc điểm số lượng của hiện tượng. Chính vì vậy, ngoài tác dụng phân tích giúp ta nhận thức được những đặc điểm cơ bản của hiện tượng bằng trực quan một cách dễ dàng và nhanh chóng, đồ thị thống kê còn là một phương pháp trình bày các thông tin thống kê một cách khái quát và sinh động, chứa đựng tính mỹ thuật, thu hút sự chú ý của người đọc, giúp người xem dễ hiểu, dễ nhớ nên có tác dụng tuyên truyền cổ động rất tốt.

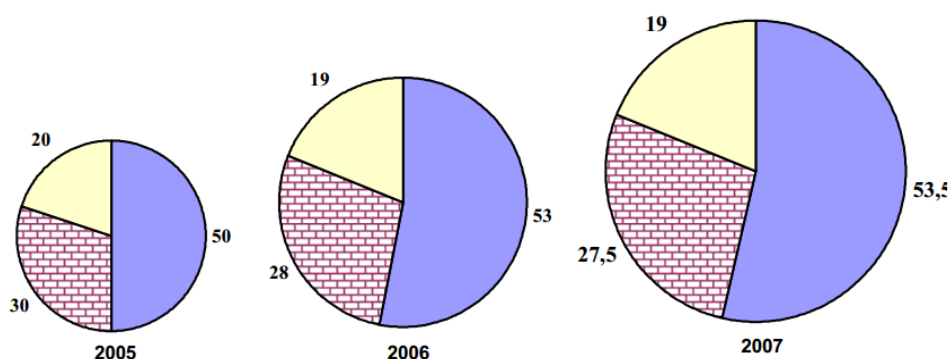
2.1. Biểu đồ hình quạt

Biểu đồ hình quạt dùng để biểu diễn dữ liệu thuộc các lớp hoặc các nhóm khác nhau bằng các miếng tỷ lệ với tần suất hoặc số lượng tương ứng. Biểu đồ dạng bánh cũng thường được sử dụng để so sánh, vì tỷ lệ dưới dạng miếng dễ quan sát hơn bằng mắt thường hơn chiều cao của từng cột. Tổng diện tích của cả phần là 100%, diện tích từng phần tương ứng với mỗi bộ phận.

Bảng : Học sinh phổ thông phân theo cấp học của địa phương X, 2005 - 2007

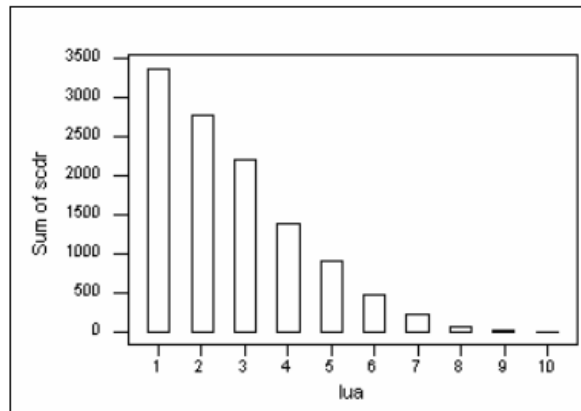
	2005		2006		2007	
	Số lượng (Người)	Cơ cấu (%)	Số lượng (Người)	Cơ cấu (%)	Số lượng (Người)	Cơ cấu (%)
Tổng số học sinh	1.000	100,0	1.140	100,0	1.310	100,0
<i>Chia ra:</i>						
Tiểu học	500	50,0	600	53,0	700	53,5
Trung học cơ sở	300	30,0	320	28,0	360	27,5
Trung học phổ thông	200	20,0	220	19,0	250	19,0

Biểu đồ cơ cấu học sinh phổ thông địa phương X từ 2005 – 2007

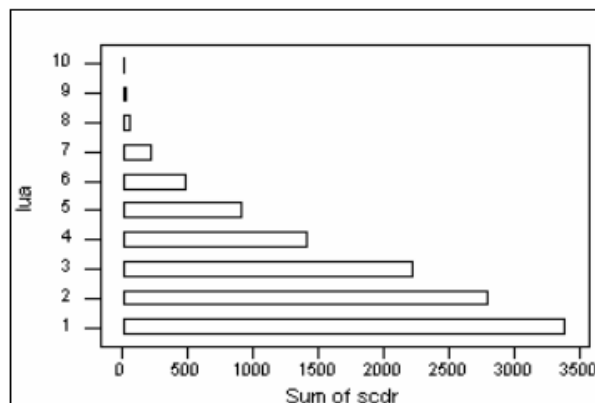


2.2. Biểu đồ hình cột

Trong biểu đồ hình cột từng nhóm trong một biến được thể hiện dưới dạng cột. Diện tích của các cột và các khoảng trống ở trục hoành đều không có ý nghĩa. Điều quan trọng của biểu đồ dạng này là chiều cao (nếu là cột thẳng đứng) hoặc chiều dài (nếu là cột nằm ngang) của các cột. Chiều cao hoặc chiều rộng sẽ tỷ lệ với phần trăm của từng nhóm.



Biểu đồ dạng cột đứng



Biểu đồ dạng cột nằm ngang

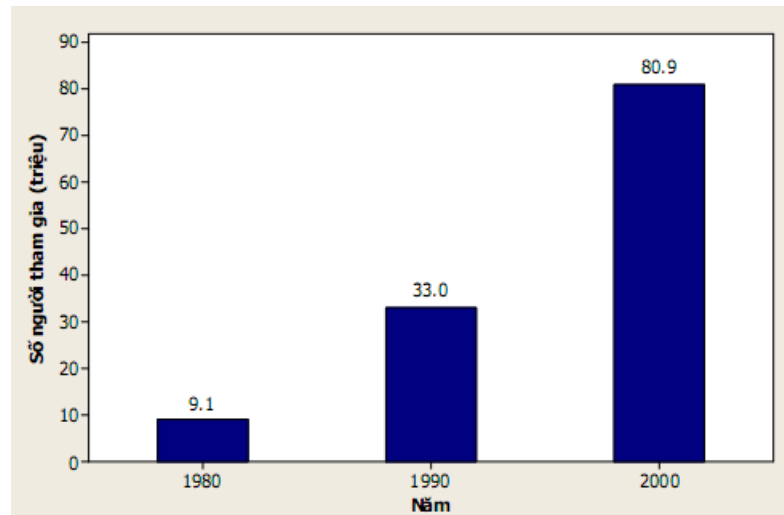
Biểu đồ tần số là đồ thị trình bày các nhóm trên trục hoành và tần số của các nhóm trên trục tung. Tần số của mỗi nhóm được tượng trưng bởi một thanh đứng mà chiều cao của thanh bằng với tần số của nhóm.

Biểu đồ tần suất là đồ thị trình bày các nhóm trên trục hoành và tần suất của các nhóm trên trục tung. Tần suất của mỗi nhóm được tiêu biểu bằng một thanh đứng mà chiều cao của thanh bằng với tần suất của nhóm.

Dùng so sánh các chuỗi dữ liệu

Thí dụ: số người tham gia tổ chức bảo vệ sức khỏe (HMOs = Health Maintenance Organization) là 9,1 triệu trong năm 1980, 33,0 triệu năm 1990 và 80,9 triệu trong năm 2000. Vẽ biểu đồ bar chart.

Graph > Bar Chart
 Bars Represent: **Values from a Table**
 Simple (two-way table) > OK
 Nhập các tùy chọn
 Graph variable: **HMO**
 Categorical variable: **Year**
 Labels... > Data labels ☒ Use labels from column: **HMO**

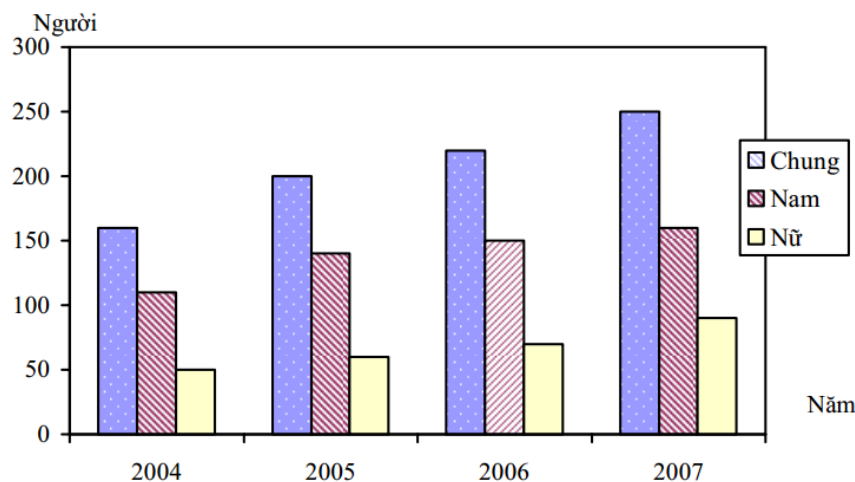


Biểu đồ hình cột là loại biểu đồ biểu hiện các tài liệu thống kê bằng các hình chữ nhật hay khối chữ nhật thẳng đứng hoặc nằm ngang có chiều rộng và chiều sâu bằng nhau, còn chiều cao tương ứng với các đại lượng cần biểu hiện.

Biểu đồ hình cột được dùng để biểu hiện quá trình phát triển, phản ánh cơ cấu và thay đổi cơ cấu hoặc so sánh cũng như biểu hiện mối liên hệ giữa các hiện tượng.

Ví dụ Biểu diễn số lượng cán bộ khoa học công nghệ của một quốc gia nào đó chia theo nam nữ của 4 năm: 2004, 2005, 2006 và 2007 qua biểu đồ

Biểu đồ : Hình cột phản ánh số lượng cán bộ khoa học công nghệ của quốc gia X, 2004 - 2007



2.3. Biểu đồ hình gấp khúc

Đồ thị đường gấp khúc là loại đồ thị thống kê biểu hiện các tài liệu bằng một đường gấp khúc nối liền các điểm trên một hệ tọa độ, thường là hệ tọa độ vuông góc. Đồ thị đường gấp khúc được dùng để biểu hiện quá trình phát triển của hiện tượng, biểu hiện tình hình phân phối, tình hình thực hiện kế hoạch theo từng tiêu chí nào đó ví dụ theo thời gian nghiên cứu.

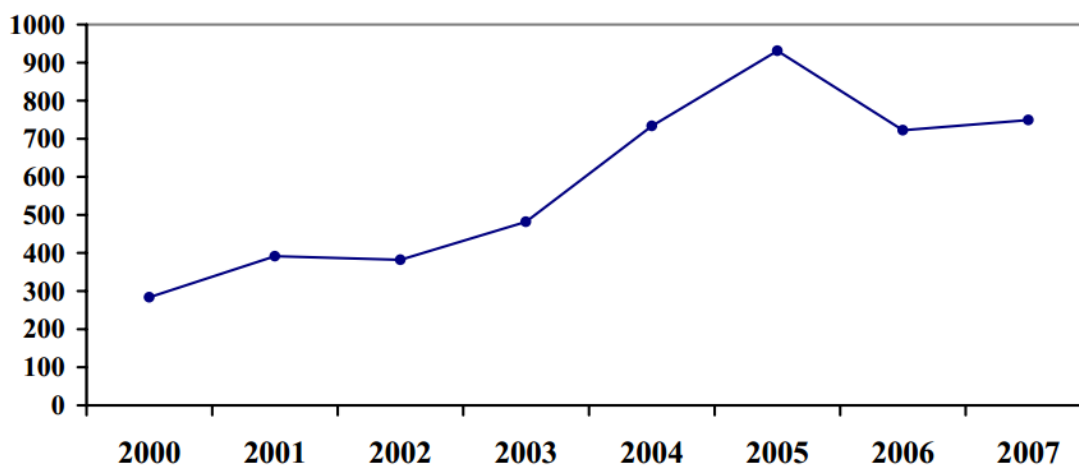
Trong đồ thị đường gấp khúc, trục hoành thường được biểu thị thời gian, trục tung biểu thị mức độ chỉ tiêu nghiên cứu.

Ví dụ : Sản lượng cà phê xuất khẩu của quốc gia X qua các năm từ 2000 - 2007 (ngàn tấn) có kết quả như sau:

Năm	2000	2001	2002	2003	2004	2005	2006	2007
Sản lượng (ngàn tấn)	283,3	391,6	382,0	482,0	733,9	931,0	722,0	749,0

Số liệu trên được biểu diễn qua đồ thị đường gấp khúc sau:

Đồ thị : Sản lượng cà phê xuất khẩu của quốc gia X từ 2000 – 2007



Thường dùng để *biểu diễn sự thay đổi của các dữ liệu theo thời gian*.

Thí dụ: Theo dõi nồng độ của Hg trong 20 năm ở hai vị trí khác nhau của Địa Trung hải. Ở 45 mẫu được thu thập ở độ sâu 10m và mang về phòng thí nghiệm để xác định nồng độ Hg. Nồng độ Hg trung bình được ghi nhận theo từng năm như trong bảng sau:

Nồng độ thủy ngân

Year	Site 1	Site 2	Year	Site 1	Site 2
1992	14.80	70.20	1982	21.50	147.8
1991	12.90	160.50	1981	18.20	197.7
1990	18.00	102.80	1980	25.80	162.1
1989	8.70	100.30	1979	11.00	123.3
1988	18.30	103.10	1978	16.50	163.9
1987	10.30	129.00	1977	28.10	129.4
1986	19.30	156.20	1976	50.50	142.6
1985	12.70	117.60	1975	60.10	169.9
1984	15.20	170.60	1974	96.70	105.1
1983	24.60	139.60	1973	100.40	162.0

Dùng dữ liệu trong bảng trên để vẽ biểu đồ line bằng chương trình Minitab:

Graph > Time Series Plot... ⇒ Multiple > OK

Nhập các tùy chọn:

Series: 'Site 1' 'Site 2'

Time/Scale... > Time Scale: ☒ **Calendar**

Start Values: ☒ **One set for each value**

Data increment: **> OK**

Data view... > Data display > ☒ **symbol**

☒ **connect line**

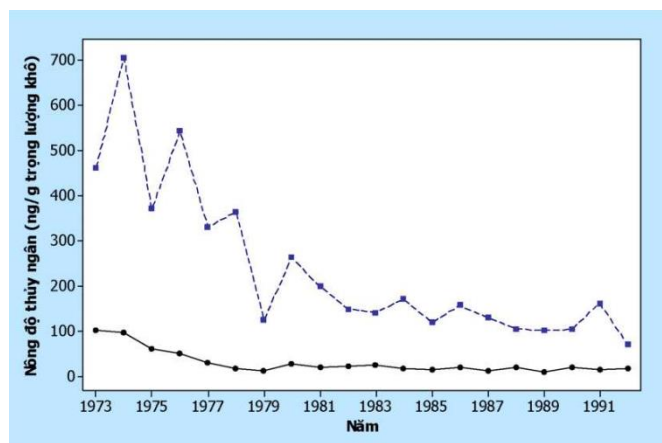
Mặc nhiên khi đưa cả hai 'site' vào cùng một biểu đồ, Minitab sẽ chỉ dùng một giá trị để làm thang đo của trục Y. Trong trường hợp hai 'site' có giá trị khác biệt nhiều (chẳng hạn trong trường hợp trên là từ 8.7 - 100.4 ở 'site 1' và 70.2 - 462 ở 'site 2') thì sẽ rất khó để thấy được biến đổi nồng độ ở 'site 1'. Trong trường hợp này việc thiết lập thêm một thang đo thứ hai cho trục Y là rất cần thiết.

Kết quả sau khi thiết lập thêm thang đo thứ hai cho trục Y.

Click chuột phải lên trục Y, chọn Edit Y scale... (Ctrl + T) > Secondary

	variable	scale	
1	site 1	secondary	T
2	site 2	primary	

OK

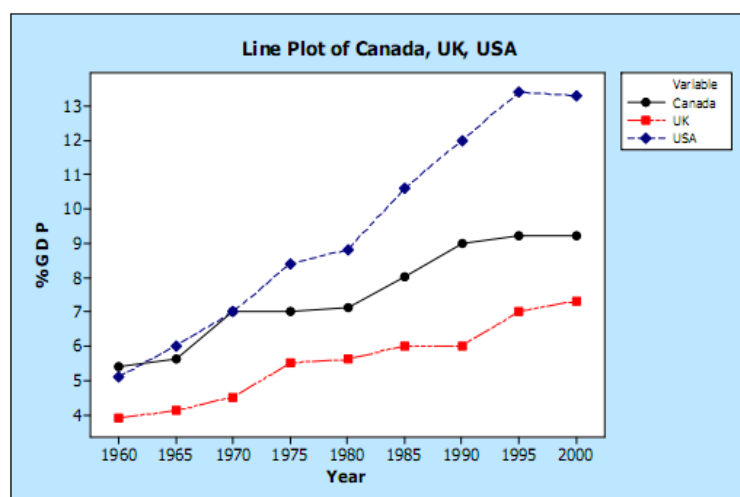


Hình: Time Series Plot của Site 1 và Site 2 với 1 trục Y

- **Biểu đồ line plots:**

Graph > Line Plots... >/With symbol > Series in Rows or Columns
 Graph variables: 'Canada' 'UK' 'USA' (C2-C4)
 Label column (optional): Year (C1)
 Series Arrangement: $\odot$ Each column forms a series

Hình biểu đồ line plots



3. Các tham số đặc trưng của mẫu

3.1. Số trung bình cộng

Số trung bình cộng, số trung bình (trung bình số học): là tổng các giá trị quan sát chia cho tổng số quan sát.

3.2. Phương sai

Còn gọi là trung bình bình phương (mean square = MS), là tham số đặc trưng tiêu biểu nhất cho tính chất phân tán của tổng thể. 1.3.2.

Giá trị trung bình của tổng thể là μ thì phương sai tổng thể là σ^2

Giá trị trung bình của mẫu là $\bar{x}$ thì phương sai mẫu là s^2

Hoặc

Còn gọi là trung bình bình phương (mean square = MS), là tham số đặc trưng tiêu biểu nhất cho tính chất phân tán của tổng thể.

3.3. Độ lệch chuẩn

Độ lệch chuẩn (Standard deviation = SD): để xác định mức độ biến động của đơn vị quan sát, ta tiến hành lấy căn bậc 2 của phương sai.

3.4. Hệ số biến thiên

Còn gọi là hệ số (biến động, Coefficient of Variation = Cv), để đo lường độ phân tán của tổng thể, được tính bằng tỷ số giữa độ lệch chuẩn và số trung bình.

3.5. Sai số trung bình

Sai số của số trung bình (Sai số chuẩn - standard error – SE)

Sai số chuẩn là tỷ số giữa độ lệch chuẩn và căn bậc hai của cỡ mẫu.

$$SE = \frac{SD}{\sqrt{n}}$$

$$s_x = \frac{s}{\sqrt{n}} \quad \text{hay} \quad s_x = \sqrt{\frac{s^2}{n}}$$

Đối với các giá trị trung bình, người ta sử dụng sai số tiêu chuẩn của giá trị trung bình thay thế cho độ lệch chuẩn s.

4. Phép thử T

4.1. Kiểm tra mức độ tin cậy của số trung bình mẫu so với số trung bình của tổng thể

- Nếu so sánh $\alpha = 0,05$, so sánh với $Z_{0,05} = 1,645$ (bảng 1 đuôi)

Hoặc $Z_{0,025} = 1,96$ (bảng 2 đuôi)

- Nếu so sánh $\alpha = 0,01$, so sánh với $Z_{0,01} = 2,33$ (bảng 1 đuôi)

Hoặc $Z_{0,005} = 2,58$ (bảng 2 đuôi)

b1: Đặt giả thiết:

- H_0 - Giá trị trung bình của mẫu bằng trung bình của quần thể ban đầu (quần thể rút mẫu). $\mu = x$
- H_1 - Giá trị trung bình của mẫu khác so với tổng thể ban đầu (quần thể rút mẫu), $\mu \neq x$

b2: Tính giá trị t hoặc z

- Tính giá trị z thực nghiệm $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$\text{hoặc tính giá trị t thực nghiệm } t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

b3: Tính giá trị P

- Xác định giá trị P: bằng cách so sánh giá trị z thực nghiệm với phân bố z.
- Xác định giá trị P: bằng cách so sánh giá trị t thực nghiệm với phân bố t.

b4: Kết luận

- Nếu $P > 0,05$ chấp nhận giả thuyết $H_0 \rightarrow Z_c > Z_\alpha$

- Nếu $P < 0,05$ ta bác bỏ H_0 tức là chấp nhận $H_1 \rightarrow Z_c < Z_\alpha$

Tóm lại: Biết độ lệch chuẩn σ , sử dụng phép thử z

Không biết độ lệch chuẩn σ , sử dụng phép thử t khi biết $\bar{x}$ và biết s

4.2. So sánh hai số trung bình mẫu với nhau

Số trung bình của mẫu ($n < 30$)

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Phương sai của mẫu:

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

- Độ lệch chuẩn của mẫu

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

- Hệ số biến động của mẫu:

$$CV\% = \frac{s}{\bar{x}} \times 100$$

với s : độ lệch chuẩn
 $\bar{x}$: trung bình mẫu

4.3. So sánh tỉ lệ trung bình quan sát với tỉ lệ trung bình lý thuyết

Khi ta khảo sát sự thay đổi tỷ lệ của một đặc tính nào đó trên mẫu do sự tác động của điều kiện nào đó bằng cách so sánh tỷ lệ này với tỷ lệ p_0 của tổng thể mà từ đó mẫu được rút ra.

Giả sử p là xác suất xuất hiện biến cố A trong một phép thử ngẫu nhiên (p chưa biết). Chúng ta muốn kiểm định giả thuyết $p = p_0$, với p_0 là một số tỷ lệ đã cho.

Nếu ta thực hiện n lần phép thử một cách độc lập và quan sát thấy biến cố A xuất hiện X lần. Xác suất của biến cố A là $\bar{p}$ à giá trị ước lượng của p .

$$\bar{p} = \frac{X}{N}$$

Để thực hiện kiểm định giả thuyết

$$\left. \begin{array}{l} H_0: p_A = p_0 \\ H_1: p_A \neq p_0 \end{array} \right\}$$

p_A : tỷ lệ ở mẫu $\bar{p}_A = \frac{A}{n}$

p_0 : tỷ lệ ở tổng thể $\rightarrow q_0 = 1 - p_0$

n : cỡ mẫu

$$Z_c = \frac{p_A - p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$$

Nếu $Z_c > Z_\alpha \rightarrow$ bác bỏ giả thuyết H_0 ,
chấp nhận H_1

Nếu $Z_c < Z_\alpha \rightarrow$ chấp nhận H_0

Bài tập:

1/ Số người đăng ký dự thi tốt nghiệp PTTH tại tỉnh A trong năm 2003 là 68842, trong đó có 32682 là học sinh nam. Hỏi tỷ lệ học sinh nam có thật sự cao hơn số nữ không?

2/ Ở một địa phương, tỷ lệ nhiễm ký sinh trùng A trên bò được xác định là 34%. Sau một thời gian dùng thuốc điều trị, lấy 100 mẫu để kiểm tra thấy có 20 bò bị bệnh. Hỏi với mức ý nghĩa $\alpha = 5\%$ hỏi thuốc này có hiệu quả không?

3/ Một nhà máy sản xuất có tỷ lệ sản phẩm đạt chất lượng là 98%. Sau một thời gian hoạt động, người ta nghi ngờ tỷ lệ trên đã bị giảm. Kiểm tra ngẫu nhiên 500 sản phẩm thấy có 28 phế phẩm. Như vậy, có thể kết luận chất lượng làm việc của máy không còn được như trước được hay không?.

4/ Tỷ lệ đẻ cá rô phi được bình thường là 50%. Xử lý thuốc chuyển giới tính trên cá rô để mong có nhiều cá đực hơn bình thường. Sau khi hoàn thành đã lấy 150 mẫu con cá để khảo sát và có 84 con cá đực. Anh chị hãy phân tích số liệu này và cho biết về hiệu quả của việc xử lý.

So sánh hai tỷ lệ quan sát (trường hợp mẫu lớn)

$$Z_c = \frac{p_A - p_B}{\sqrt{p \cdot q \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Với $p = \frac{x_A + x_B}{n_A + n_B}$, $x_A = n_A \cdot p_A$ và $x_B = n_B \cdot p_B$