

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



**GIÁO TRÌNH
MÔN HỌC: PHƯƠNG PHÁP THÍ NGHIỆM
NGÀNH, NGHỀ: THÚ Y
TRÌNH ĐỘ: TRUNG CẤP/CAO ĐẲNG**

(Ban hành kèm theo Quyết định số 257/QĐ-TCĐNĐT-ĐT ngày 13 tháng 07 năm 2017
của Hiệu trưởng trường Cao đẳng Nghề Đồng Tháp)

Đồng Tháp, năm 2017

CHƯƠNG 1. NHỮNG VẤN ĐỀ CƠ BẢN CỦA THỐNG KÊ SINH HỌC

1.1. Một số khái niệm cơ bản.

1.1.1. Tổng thể ($n \geq 30$)

- 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 có thể vô hạn hoặc hữu hạn. Với những tổng thể vô hạn, 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.

- 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.

1.1.2. Mẫu ($n < 30$)

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

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.

1.1.3. Số liệu, Dữ liệu (Data): 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...

a/ 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...

b/ 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...

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

1.1.4. 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.

1.2. Các dạng biểu đồ thường gặp.

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 các 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.

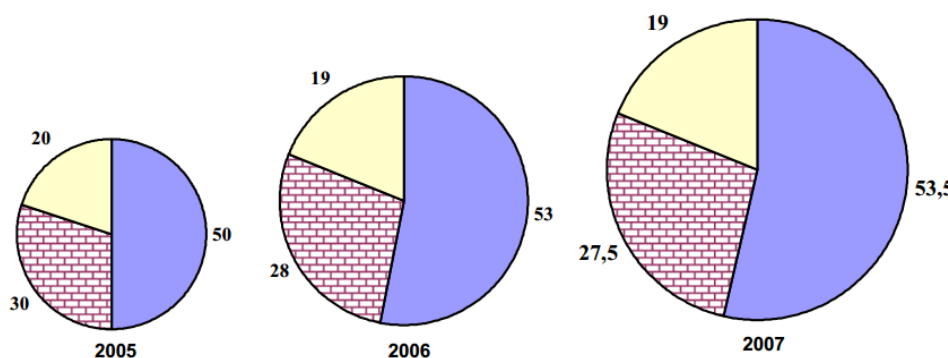
1.2.1. Biểu đồ hình quạt (dạng bánh)

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



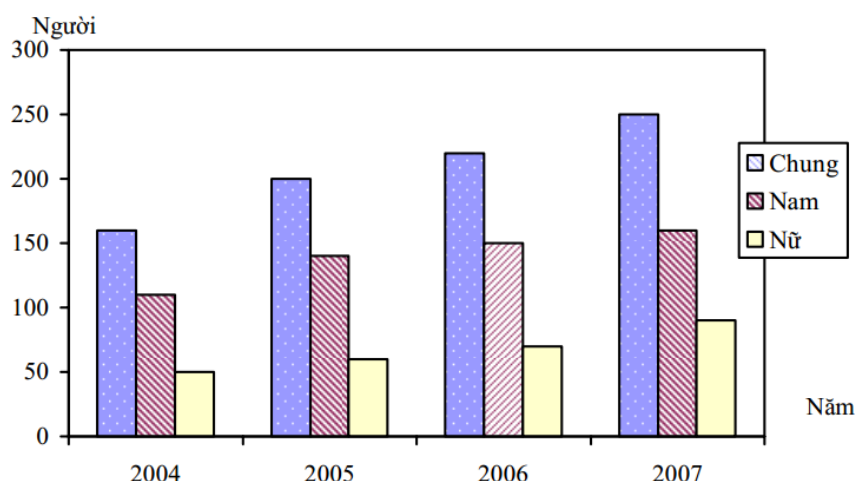
1.2.2. Biểu đồ hình cột

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



1.2.3. 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.

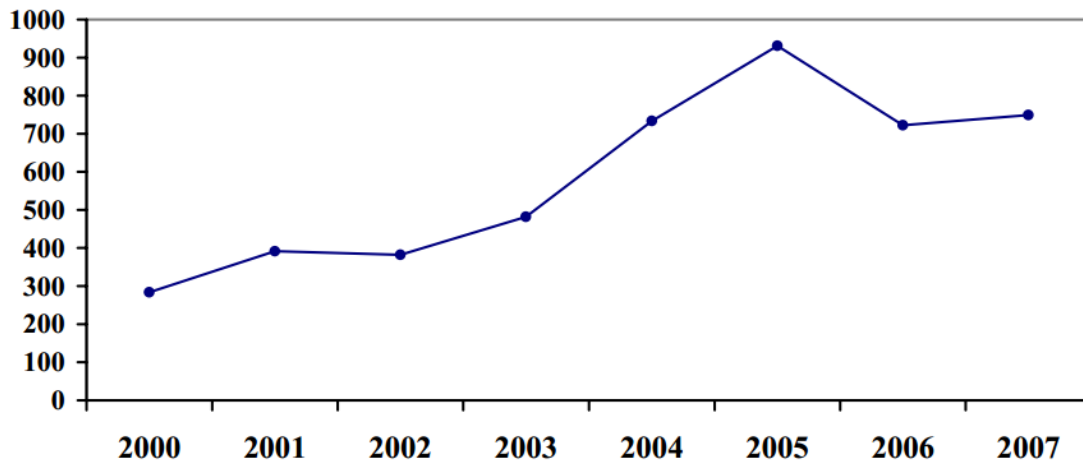
- Trong đồ thị đường gấp khúc, trục hoành thường được biểu thị thời gian, trục tung biểu thị mức độ chỉ tiêu nghiên cứu.

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



1.3 Các tham số đặc trưng đo lường khuynh hướng tập trung của các giá trị

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

1.3.2 Số mốt (mode): là giá trị có tần suất cao nhất trong bộ dữ liệu. Trong phân bố tần suất, mode là giá trị nằm ở điểm cao nhất trên đường cong. Đối với phân phối chuẩn thì mode chính là trung vị và trung bình.

1.3.4 Số trung vị (Median): là giá trị nằm giữa bộ số liệu. Lợi ích của trung vị là khi dữ liệu chứa các giá trị rất lớn với tần số thấp chúng sẽ ảnh hưởng mạnh đến trung bình số học, nhưng không ảnh hưởng đến giá trị trung vị. Do đó lúc này trung vị cho ta một ý niệm tốt hơn về giá trị trung tâm của phân phối. Sắp xếp số liệu theo thứ tự tăng dần, đánh số thứ tự cho các dữ liệu.

$$M = (n+1)/2$$

n: số thứ tự lớn nhất sau khi đã được đánh số

1.4 Các tham số đo lường sự phân tán của các giá trị

1.4.1 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

1.4.2. 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.

1.4.3. Hệ số biến dị (biến động, biến thiên- Coefficient of Variation = Cv):

Hệ số biến thiên là chỉ tiêu tương đối phản ánh mối quan hệ so sánh giữa độ lệch chuẩn và số bình quân số học để đo lường độ phân tán của tổng thể, được tính bằng tỷ số giữa độ lệch chuẩn và số trung bình.

1.4.4. 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.

1.5 Công thức tính các tham số:

1.5.1 Tính từ biến quan sát theo dạng chuỗi dữ liệu

a/ Tổng thể ($n \geq 30$),

- Số trung bình của tổng thể (population mean)

$$\mu = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N}$$

- Phương sai của tổng thể:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

- Độ lệch chuẩn của tổng thể: **SD**

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{(x_i - \mu)^2}{N}}$$

- Hệ số biến động của tổng thể

$$CV \% = \frac{\sigma}{\mu} \times 100$$

với σ : độ lệch chuẩn
 μ : trung bình

b/ Mẫu ($n < 30$)

- Số trung bình của mẫu:

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

- Phương sai của mẫu:

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

- Độ lệch chuẩn của mẫu: **SD**

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

- Hệ số biến động của mẫu: $CV(\%) = (SD/x \text{ trung bình}) * 100$

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

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

Bài tập 1: Hãy tính các tham số đặc trưng của thống kê trong một đàn heo 30 con, có trọng lượng như sau

STT	Trọng lượng (kg)
1	20
2	17
3	19
4	23
5	21
6	17
7	16
8	22
9	21
10	23
11	15
12	17
13	21
14	19
15	19

STT	Trọng lượng (kg)
16	22
17	19
18	20
19	18
20	20
21	20
22	22
23	20
24	18
25	18
26	16
27	22
28	24
29	22
30	20

Kết quả: Số trung bình: 19,7; Độ lệch chuẩn: 2,26; Sai số chuẩn: 11,5 ; Phương sai: 5,14

1.5.2 Tính từ dạng bảng tần số

Tần số là số lần xuất hiện của mỗi giá trị x_i trong bảng số liệu.

Ví dụ 1: Trọng lượng (kg/con) của 16 lô heo được ghi nhận như sau:

15	21	20	19	22	21	16	19
----	----	----	----	----	----	----	----

20	16	17	24	16	21	15	22
----	----	----	----	----	----	----	----

Số liệu bảng trọng lượng heo được trình bày dưới dạng bảng thực nghiệm như sau:

Trọng lượng (kg): (x_i)	Số heo (Tần số: n_i)	(Tần suất: p_i) $p_i = (n_i/n)$
15	2	0,1250
16	3	0,1875
17	1	0,0625
19	2	0,1250
20	2	0,1250
21	3	0,1875
22	2	0,1250
24	1	0,0625
Tổng	16	1,000

- Số trung bình của giá trị dạng tần số

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

- Phương sai của dạng bảng tần số

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

Ví dụ 1: Trọng lượng của 6 heo con được ghi nhận như sau:

Trọng lượng (kg) x_i	Số heo (tần số: n_i)	Tần suất ($p_i = (n_i/n)$)
15	2	0.33
16	3	0.5
17	1	0.17
Tổng	6	1.000

Tính trung bình tổng thể dạng bảng tần số. *Đáp án* $\mu = 15,83\text{kg}$.

Tính trung bình tổng thể dạng bảng tần suất. *Đáp án* $\mu = 15,84\text{kg}$.

1.5.3 Tính từ dạng bảng tần suất:

Tần suất là f_i của giá trị x_i là tỷ số giữa tần số n_i và kích thước mẫu n

$$p_i = \frac{n_i}{n}$$

- Số trung bình của giá trị dạng tần suất: (trung bình tổng thể và trung bình mẫu)

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

- Phương sai của dạng bảng tần suất:

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

Ví dụ 2: Trọng lượng (kg) của một đàn heo có 5 con như sau: 20, 17, 19, 23, 21. Hãy tính trọng lượng trung bình của đàn heo.

Giải: với $n=5$. Số trung bình của giá trị biến quan sát (hay dạng chuỗi dữ liệu):

$$\text{Mẫu: } (n < 30) \quad \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \rightarrow \bar{x} = 20 \text{ kg.}$$

Ví dụ 3: Trọng lượng của 6 heo con được ghi nhận như sau:

Trọng lượng (kg) x_i	Số heo (tần số n_i)	Tần suất ($f_i = n_i/n$)
15	2	0.33
16	3	0.5
17	1	0.17
Tổng	6	1.000

Tính trung bình tổng thể dạng bảng tần số. *Đáp án* $\mu = 15,83\text{kg}$.

Tính trung bình tổng thể dạng bảng tần suất. *Đáp án* $\mu = 15,84\text{kg}$.

Ví dụ 4: Đánh giá trọng lượng của một lô gà ở trại A, có khối lượng (kg) như sau:

2.45	4.0	3.25	3.75	4.05
------	-----	------	------	------

Đáp án: $\bar{x}$: 3.5 (kg), $s^2 = 0.445$, $s = 0.667$, $CV\% = 19.06$

1.6 Kiểm định giả thuyết

Các đặc trưng của mẫu ngoài việc dùng để ước lượng các đặc trưng của tổng thể còn được dùng để đánh giá xem một giả thuyết nào đó của tổng thể là đúng hay sai. Việc tìm ra kết luận để bác bỏ hay chấp nhận một giả thuyết được gọi là kiểm định giả thuyết.

Giả thuyết đưa ra kiểm định được ký hiệu là H_0 được gọi là giả thuyết không. Đây là giả thuyết mà ta nghi ngờ và muốn bác bỏ.

Ngoài ra còn một giả thuyết nữa gọi là giả thuyết đối, ký hiệu là H_1 . H_1 sẽ được chấp nhận khi H_0 bị bác bỏ.

Chúng ta bác bỏ hay chấp nhận một giả thuyết bằng cách nào?. Các nhà thống kê đều nhất trí nguyên lý sau: “Nếu một biến cố có xác suất rất nhỏ thì trong một phép thử biến cố đó sẽ không xảy ra”. Như vậy chúng ta sẽ quyết định bác bỏ giả thuyết H_0 nếu xác suất xuất hiện của một sự kiện quan sát được là “nhỏ”.

Thực tế \ Kết luận	Chấp nhận H_0	Bác bỏ H_0
H_0 đúng	Kết luận đúng	Sai lầm loại I
H_0 sai	Sai lầm loại II	Kết luận đúng

Các bước để kiểm định giả thuyết:

Bước 1: Đặt giả thuyết.

Có 3 dạng giả thuyết:

$H_0 = \mu = \mu_0$ $H_1 = \mu \neq \mu_0$ Kết quả: dạng số liệu nằm trong khoảng $a >$ hoặc $< b$.	$H_0 = \mu = \mu_0$ $H_1 = \mu > \mu_0$ Kết quả: $>$ hơn không đạt yêu cầu. Ví dụ kiểm tra hàm lượng chất độc hại.	$H_0 = \mu = \mu_0$ $H_1 = \mu < \mu_0$ Kết quả: $<$ hơn không đạt yêu cầu. Ví dụ kiểm tra hàm lượng dinh dưỡng cần thiết.
--	--	--

Bước 2: lấy mẫu, chọn số thống kê tương ứng với giả thuyết.

Chọn công thức chứa các giá trị của mẫu

Là biến ngẫu nhiên để biết được phân phối của xác suất

Giá trị của nó liên quan đến xác suất để giả thuyết đúng.

Bước 3: Thế các giá trị của mẫu vào để tính các giá trị của nó