

THỐNG KÊ MÔ TẢ

Chương 3
TRÌNH BÀY DỮ LIỆU

- 1. Phân tổ
- 2. Bảng tần số
- 3. Bảng tần số kết hợp
- 4. Biểu đồ, đồ thị

1. Phân tổ

- 2. Nguyên tắc:
 - ✓ Những cá thể xếp cùng tổ / nhóm có tính chất càng tương tự nhau càng tốt.
 - ✓ Không thừa tổ: tổ **rỗng**, không chứa cá thể nào.
 - ✓ Không thiếu tổ: có cá thể không thuộc được vào tổ nào.

1. Phân tổ

- 3. Các bước phân tổ.
 - Bước 2:** Xác định số tổ
 - TH2: Biến có **nhều** giá trị
- Ghép những giá trị tương tự (gần giống) nhau làm cơ sở hình thành một tổ (đa trị).

1. Phân tổ

- 1. Khái niệm:

Là việc căn cứ vào một hay một số biến nhất định để phân chia tập dữ liệu (tổng thể hay mẫu) thành các tổ / nhóm khác nhau và bố trí và sắp xếp các cá thể trong tập dữ liệu vào tổ / nhóm.

Ý nghĩa: Là phương pháp cơ bản và quan trọng tiến hành tổng hợp và phân tích thống kê.

1. Phân tổ

- 3. Các bước phân tổ.
 - Bước 1:** Lựa chọn biến phân tổ
 - Là việc xác định biến được chọn làm căn cứ để phân tổ.
 - Bước 2:** Xác định số tổ
 - TH1: Biến có **ít** giá trị
- Mỗi giá trị làm cơ sở hình thành một tổ (đơn trị).

1. Phân tổ

- 4. Tổ đa trị định lượng.

Khi biến phân tổ là biến định lượng → hình thành các **tổ đa trị định lượng.**

 - a) Vùng trị số:

Mỗi tổ sẽ là một vùng trị số giữa hai giá trị gọi là **giới hạn dưới** và **giới hạn trên** của tổ.

Các tổ được sắp xếp theo thứ tự vùng trị số tăng dần.

1. Phân tổ

4. Tổ đa trị định lượng.
- a) Vùng trị số:
Biến liên tục: giới hạn trên tổ trước = giới hạn dưới tổ sau
Biến rời rạc: giới hạn dưới tổ sau = giới hạn trên tổ trước + 1 đơn vị tính
- b) **Khoảng cách tổ:** là độ lệch giữa giới hạn trên và giới hạn dưới của tổ, $h = x_{\max} - x_{\min}$

1. Phân tổ

4. Tổ đa trị định lượng.
- d) Phân tổ đều
+ Là phân thành các tổ có khoảng cách bằng nhau.
+ Số tổ đều:
- $$k = \sqrt[3]{2n}$$

n: số cá thể của tập dữ liệu

1. Phân tổ

3. Tổ đa trị định lượng.
- d) Phân tổ đều
Số tổ đều k và khoảng cách tổ đều h thường lẻ nên được làm tròn thành hai số tự nhiên gần nhất nhưng phải đảm bảo không **thiếu** tổ, **thừa** tổ (rỗng).

1. Phân tổ

4. Tổ đa trị định lượng.
- c) Tổ mở: là tổ thiếu giới hạn dưới hoặc giới hạn trên, xuất hiện ở tổ đầu tiên hoặc cuối cùng.
Ví dụ: Điểm học tập trung bình của sinh viên chia thành:
> 9 : Xuất sắc, 8 – 9 : Giỏi, 7 – 8 : Khá, 5 – 7 : Trung bình, 3 – 5 : Yếu, < 3 : Kém

1. Phân tổ

3. Tổ đa trị định lượng.
- d) Phân tổ đều
+ Khoảng cách tổ đều:
Biến liên tục:
- $$h = \frac{x_{\max} - x_{\min}}{k}$$

Biến rời rạc:

$$h = \frac{x_{\max} - x_{\min} - (k - 1)}{k}$$

1. Phân tổ

- Ví dụ: Số liệu về giá trị xuất khẩu của 30 doanh nghiệp công nghiệp Hà nội năm 2005 (triệu/USD) như sau. Hãy phân tổ đều dữ liệu này.
- | | | | | | |
|----|----|----|----|----|----|
| 27 | 49 | 34 | 40 | 50 | 25 |
| 34 | 20 | 30 | 35 | 45 | 28 |
| 46 | 30 | 25 | 58 | 25 | 38 |
| 26 | 28 | 36 | 32 | 24 | 36 |
| 60 | 25 | 33 | 28 | 46 | 25 |

2. Bảng tần số

1. Khái niệm:
- a) Tần số (f):
- Là số lần lặp lại (xuất hiện) trong một khoảng xác định.
 - Ý nghĩa: cho biết số cá thể của tổ trong tập dữ liệu.
- b) Tần suất (tần số tương đối):
- Là tỉ lệ (%) giữa tần số với cỡ mẫu
 - Ý nghĩa: Cho biết tỷ trọng của tổ trong tập dữ liệu.

2. Bảng tần số

1. Khái niệm:
- c) Tần số tích lũy:
- Ý nghĩa:
 - + Tổ đơn trị (không có khoảng cách tổ): cho biết số cá thể có trị số nhỏ hơn hoặc bằng trị số tổ.
 - + Tổ đa trị (có khoảng cách tổ): phản ánh số cá thể có giá trị nhỏ hơn giới hạn trên của tổ.
- d) Tần suất tích lũy (tần số tích lũy tương đối):
- Là tỉ lệ (%) giữa tần số tích lũy với cỡ mẫu.

2. Bảng tần số

2. Cấu tạo:
- Bảng tần số gồm 2 thành phần:
- Các giá trị hoặc vùng giá trị của biến phân tổ.
 - Tần số của tổ và những thông tin cần thiết khác như tần số tích lũy, tần suất, tần suất tích lũy, mật độ phân phối tương ứng của tổ.

2. Bảng tần số

1. Khái niệm:
- c) Tần số tích lũy:
- Giá trị dữ liệu sắp xếp theo thứ tự (tăng dần).
 - **Tần số tích lũy (lũy tiến)** của một tổ là tổng các tần số từ tổ đầu tiên đến tần số của tổ đó.
 - Kí hiệu: S
 - Công thức: $S_k = \sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k$

2. Bảng tần số

1. Khái niệm:
- e) Mật độ phân phối:
- Là tỉ số giữa tần số (hoặc tần suất) với khoảng cách tổ (chỉ có với tổ có khoảng cách).
 - Công thức: $d_i = \frac{f_i}{h_i}$

2. Bảng tần số

3. Ví dụ:
- | Năng suất lúa (tạ/ha) | Số hộ gia đình |
|-----------------------|----------------|
| [30 - 35) | 5 |
| [35 - 40) | 10 |
| [40 - 50) | 20 |
| [50 - 60) | 12 |
| [60 - 75) | 3 |
| Tổng | 50 |

3. Bảng tần số kết hợp (bảng chéo)

1. Khái niệm:

Là bảng trình bày các thông tin thống kê một cách có hệ thống, hợp lý, rõ ràng các **tần số** của **số nhiều (>1) biến** cùng lúc nhằm nêu lên những đặc trưng về mặt lượng của hiện tượng nghiên cứu.

3. Bảng tần số kết hợp (bảng chéo)

2. Yêu cầu khi xây dựng bảng:

- Cách ghi số liệu : Các ô trong bảng dùng để ghi số liệu, nhưng nếu không có số liệu thì dùng các kí hiệu qui ước sau:
 - + Dấu gạch ngang (-) : Hiện tượng không có số liệu.
 - + Dấu ba chấm (...) : Số liệu còn thiếu, sau này có thể bổ sung.
 - + Dấu gạch chéo (x) : Hiện tượng không liên quan đến chỉ tiêu, nếu viết số liệu vào ô đó sẽ không có ý nghĩa.

3. Bảng tần số kết hợp (bảng chéo)

2. Yêu cầu khi xây dựng bảng:

- Các tiêu đề, tiêu mục ghi chính xác, gọn, đầy đủ, dễ hiểu.
- Các chỉ tiêu giải thích cần sắp xếp hợp lý, phù hợp với mục tiêu nghiên cứu. Các chỉ tiêu có liên hệ với nhau nên sắp xếp gần nhau.
- Có đơn vị tính cụ thể cho từng chỉ tiêu.

3. Bảng tần số kết hợp (bảng chéo)

3. Bảng 2 biến định tính:

Công việc chủ hộ	Tân Bình		Quận 6		Quận 11	
	Tần số	Tần suất	Tần số	Tần suất	Tần số	Tần suất
Làm việc nhà máy	18	14.3%	91	15.6%	5	7.0%
Làm nghề tự do	39	31.0%	187	32.1%	11	15.5%
Làm việc CQNN	2	1.6%	30	5.2%	2	2.8%
Làm việc văn phòng	3	2.4%	22	3.8%	1	1.4%
Buôn bán nhỏ	4	3.2%	2	0.3%	1	1.4%
Bán hàng rong	5	4.0%	39	6.7%	1	1.4%
Làm việc tại nhà	10	7.9%	45	7.7%	7	9.9%
Tự kinh doanh	8	6.3%	88	15.1%	28	39.4%
Thu nhập tự nguồn khác	6	4.8%	37	6.4%		0.0%
Không việc làm	31	24.6%	41	7.0%	15	21.1%
Tổng	126		582		71	

3. Bảng tần số kết hợp (bảng chéo)

4. Bảng 3 biến định tính:

Công việc chủ hộ	Tân Bình			Quận 6			Quận 11		
	Nữ	Nam	Cộng	Nữ	Nam	Cộng	Nữ	Nam	Cộng
Làm việc nhà máy	3	15	18	30	61	91	2	3	5
Làm nghề tự do	12	27	39	44	143	187	4	7	11
Làm việc CQNN	1	1	2	5	25	30	2		2
Làm việc văn phòng	2	1	3	10	12	22	1		1
Buôn bán nhỏ	2	2	4	2		2		1	1
Bán hàng rong	3	2	5	20	19	39	1		1
Làm việc tại nhà	6	4	10	32	13	45	3	4	7
Tự kinh doanh	5	3	8	53	35	88	13	15	28
Thu nhập tự nguồn khác	2	4	6	8	29	37			
Không việc làm	18	13	31	25	16	41	9	6	15
Tổng	54	72	126	229	353	582	35	36	71

3. Bảng tần số kết hợp (bảng chéo)

5. Bảng biến định lượng và biến định tính:

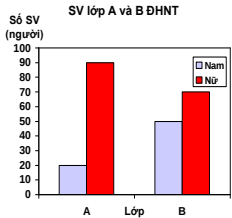
Thu nhập (ngàn đồng)	Tân Bình		Quận 6		Quận 11	
	Tần số	Tần suất	Tần số	Tần suất	Tần số	Tần suất
1 - 250	18	14.3%	91	15.6%	5	7.0%
251 - 500	39	31.0%	187	32.1%	11	15.5%
501 - 750	2	1.6%	30	5.2%	2	2.8%
751 - 1000	3	2.4%	22	3.8%	1	1.4%
1001 - 1250	4	3.2%	2	0.3%	1	1.4%
1251 - 1500	5	4.0%	39	6.7%	1	1.4%
1501 - 1750	10	7.9%	45	7.7%	7	9.9%
1751 - 2000	8	6.3%	88	15.1%	28	39.4%
2001 - 3000	6	4.8%	37	6.4%		0.0%
3001 - 4000	31	24.6%	41	7.0%	15	21.1%
Tổng	126		582		71	

4. Biểu đồ, đồ thị

1. Khái niệm:
Là các hình vẽ hoặc đường nét hình học dùng để miêu tả có tính chất qui ước các số liệu thống kê.
2. Ý nghĩa:
Ứng dụng rộng rãi trong mọi công tác nhằm hình tượng hoá về hiện tượng nghiên cứu

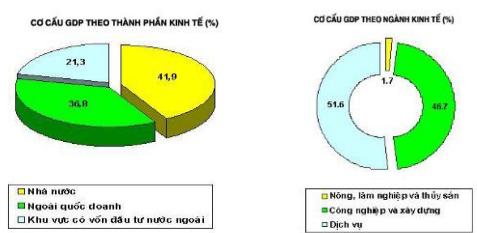
4. Biểu đồ, đồ thị

3. Các loại đồ thị:
- Căn cứ vào hình thức biểu hiện:
+ Biểu đồ thanh / cột



4. Biểu đồ, đồ thị

3. Các loại đồ thị:
+ Biểu đồ diện tích (hình vuông, hình tròn, hình chữ nhật...)

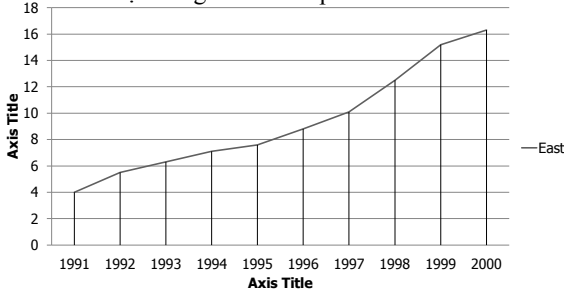


4. Biểu đồ, đồ thị

3. Các loại đồ thị:
- Căn cứ theo nội dung phản ánh:
+ Đồ thị phát triển
+ Đồ thị kết cấu
+ Đồ thị liên hệ
+ Đồ thị so sánh
+ Đồ thị phân phối
+ Đồ thị kế hoạch
.....

4. Biểu đồ, đồ thị

3. Các loại đồ thị:
+ Đồ thị đường nối liên tiếp



4. Biểu đồ, đồ thị

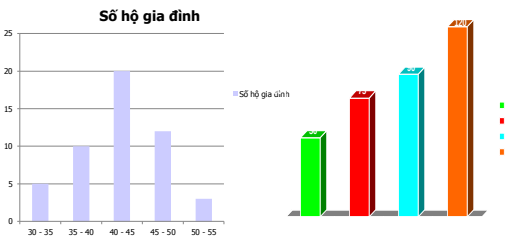
3. Các loại đồ thị:
- Căn cứ vào hình thức biểu hiện:
+ Biểu đồ tượng hình (biểu hiện bằng các hình vẽ tượng trưng)
+ Bản đồ thống kê

4. Biểu đồ, đồ thị

4. Đồ thị phân phối tần số (Histogram):
- Thường được sử dụng biểu diễn dữ liệu trên bảng phân phối tần số, trình bày dưới dạng biểu đồ hình cột.
 - Ý nghĩa:
 - + Cảm nhận tương đối về sự tập trung dữ liệu
 - + Thấy mức độ phân tán tương đối của dữ liệu
 - + Cảm nhận sơ bộ hình dáng phân phối dữ liệu

4. Biểu đồ, đồ thị

4. Đồ thị phân phối tần số (Histogram):
- Ví dụ:



4. Biểu đồ, đồ thị

5. Biểu đồ cành lá:
- Mỗi số liệu được chia thành 2 phần: phần cành và phần lá:
- + Phần cành biểu diễn giá trị của tổ.
 - + Phần lá giúp xác định tần số tổ.
- Ví dụ: Dãy số liệu : 21 ; 24; 26; 27 ; 27 ; 30 ; 32 ; 41



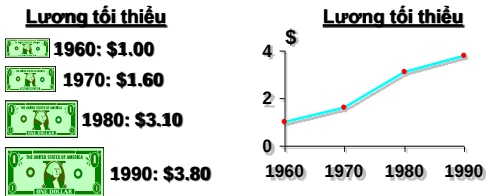
Cành	Lá
2	1 4 6 7 7
3	0 2
4	1

4. Biểu đồ, đồ thị

6. Lưu ý khi xây dựng biểu đồ:
- Lựa chọn đồ thị phù hợp với nội dung, tính chất dữ liệu.
 - Tỷ lệ xích theo các khoảng cách thích hợp, chính xác và thống nhất.
 - Ghi đầy đủ dữ liệu, đơn vị tính, thời gian, không gian thích hợp từng loại đồ thị.

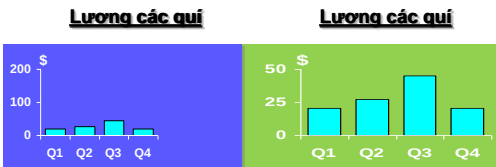
4. Biểu đồ, đồ thị

Hãy nhận xét về cách trình bày sau, cách nào tốt hơn?



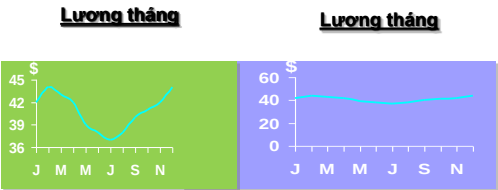
4. Biểu đồ, đồ thị

Hãy nhận xét về cách trình bày sau, cách nào tốt hơn?



4. Biểu đồ, đồ thị

Hãy nhận xét về cách trình bày sau, cách nào tốt hơn?



5. Bài tập ôn chương

a) Doanh thu các ngày của một cửa hàng internet tháng 6/2005 như sau:

Đ/v : 1000đ

550	400	520	520	700	490
450	550	400	350	540	500
660	360	550	600	300	550
460	450	580	380	490	500
590	360	560	550	550	490

5. Bài tập ôn chương

- 1/ Hãy:
- Lập biểu đồ cành lá với hàng trăm của số liệu là cành.
 - Phân số liệu thành 4 tổ đều và lập bảng tần số.
 - Biểu diễn tần số của số liệu đã phân tổ bằng đồ thị.
- của bảng dữ liệu sau:

5. Bài tập ôn chương

b) Thời gian đi làm thêm của 50 sinh viên đại học trong một tuần. Đơn vị: giờ/tuần.

2	10	3	5	1	19	14	3	15	12
30	10	2	8	19	17	12	29	21	29
15	23	3	24	4	16	49	15	64	5
3	22	19	1	48	4	25	7	22	4
10	3	33	12	14	30	10	13	5	23

5. Bài tập ôn chương

- 2/ Tập dữ liệu thô của một mẫu, gồm các biến sau:
- Biến 1: Giới tính có giá trị: 1. Nữ; 2. Nam
 - Biến 2: Tuổi
 - Biến 3: Mức độ hài lòng của người dân khi giao tiếp nơi công sở với 3 giá trị: 1. Hài Lòng; 2. Bình thường; 3. Không hài lòng
- Hãy thực hiện:
- Lập bảng tần số cho biến **Giới tính**
 - Lập bảng tần số kết hợp mô tả sự liên hệ giữa **Mức độ hài lòng của người dân khi giao tiếp nơi công sở** với **Giới tính**.

5. Bài tập ôn chương

STT	Biến 1	Biến 2	Biến 3	STT	Biến 1	Biến 2	Biến 3
1	1	32	2	9	1	32	1
2	1	36	1	10	2	33	2
3	2	24	2	11	1	28	3
4	2	37	3	12	1	38	2
5	2	27	1	13	2	31	2
6	1	32	2	14	2	35	1
7	2	32	2	15	2	26	2
8	2	31	3	16	2	31	1