

Chương II

Tổng hợp thống kê

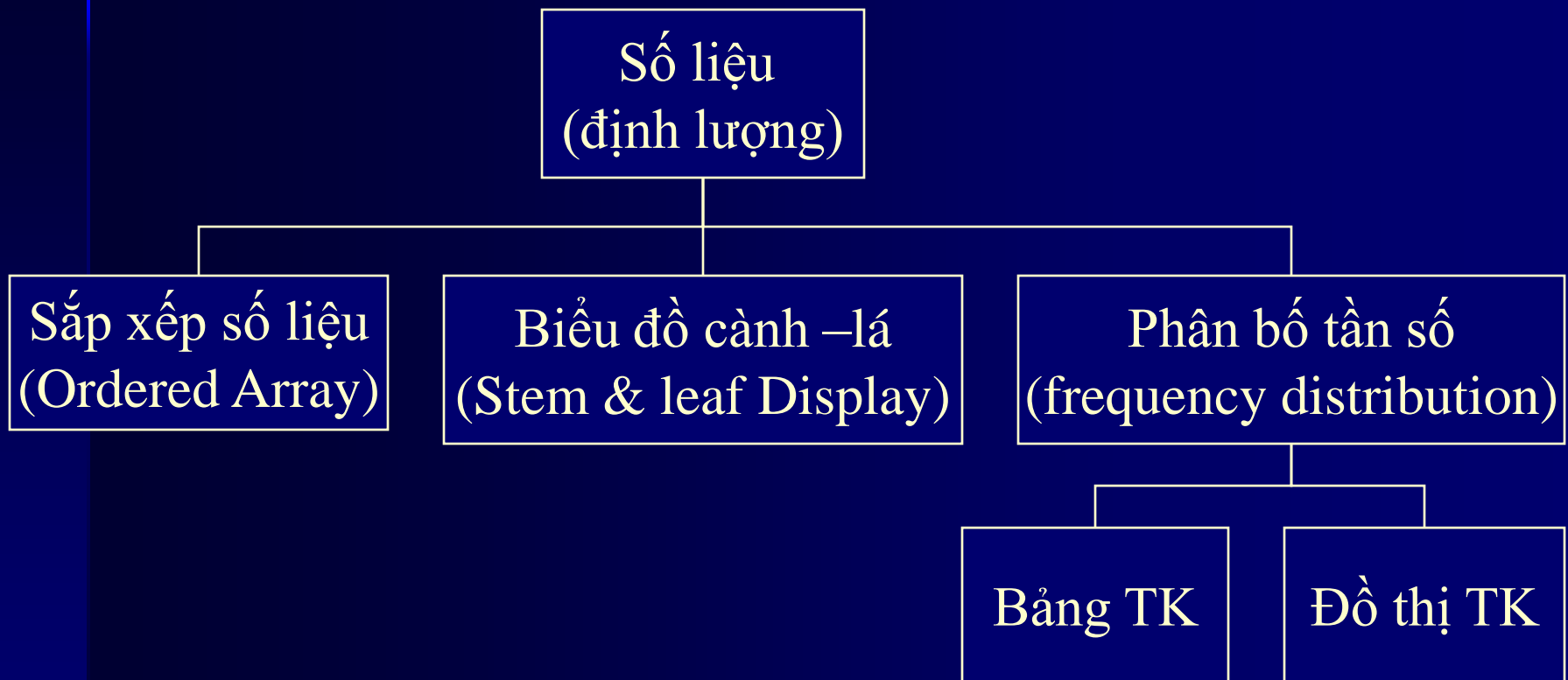
VD1

- Giám đốc một công ty tin học dự định trả mức lương 2800000 VND/tháng cho một lập trình viên làm tại công ty với 3 năm kinh nghiệm. Để biết mức lương này đã thoả đáng chưa, ông ta tổ chức một cuộc điều tra 30 lập trình viên làm cho các công ty cạnh tranh với 3 năm kinh nghiệm. Kết quả điều tra như sau:

Đ/v : 1000đ/tháng

2400	2700	2350	2900	2500	2800
2800	2200	2800	2700	2400	3000
2950	2600	2700	2300	2700	2500
2600	2300	2500	2750	2700	2750
3000	2550	2700	2350	2650	2450

Một số phương pháp tổng hợp thống kê



I - Sắp xếp số liệu (đối với số liệu định lượng)

- Cách sắp xếp
 - + Sắp xếp theo thứ tự (từ thấp đến cao hoặc ngược lại).
 - + Sắp xếp theo tính chất quan trọng.

.....

(Số liệu định tính : Sắp xếp theo trật tự văn
A,B,C; theo t/c quan trọng...)

■ VD1 :Số liệu sau khi đã được sắp xếp

Đ/v :1000đ/tháng

2200	2400	2500	2700	2700	2800
2300	2400	2550	2700	2750	2900
2300	2450	2600	2700	2750	2950
2350	2500	2600	2700	2800	3000
2350	2500	2650	2700	2800	3000

I - Sắp xếp số liệu (đối với số liệu định lượng)

- Tác dụng:
 - + Nhanh chóng phát hiện giá trị cao nhất và thấp nhất trong tập hợp số liệu.
 - + Dễ dàng chia số liệu thành nhóm
 - + Phát hiện nhanh giá trị nào xuất hiện bao nhiêu lần
 - + Quan sát khoảng cách giữa các số liệu liên tiếp nhau
- Hạn chế : Không thích hợp với lượng thông tin quá lớn.

II - Biểu đồ cành lá (dùng đối với số liệu định lượng)

- Mỗi số liệu được chia thành 2 phần : phần thân và phần lá:

+ Phần thân xác định thứ bậc

+ Phần lá dùng để xác định tần số (đếm)

VD: Dãy số liệu : 21 ; 24; 26 ; 27 ; 27 ; 30 ; 32 ; 41

2		14677
3		02
4		1

II - Biểu đồ cành lá (dùng đối với số liệu định lượng)

VD1 : Có thể sử dụng 2 số đầu là thân, 2 số sau là lá :

22	00
23	00 00 50 50
24	00 00 50
25	00 00 00 50
26	00 00 50
27	00 00 00 00 00 00 50 50
28	00 00 00
29	00 50
30	00 00

KL

III – Phân tổ thống kê

1/ KN, ý nghĩa, nhiệm vụ của phân tổ thống kê

a- KN :

Là việc phân chia các đơn vị của tổng thể thống kê thành các tổ (và các tiểu tổ) có tính chất khác nhau trên cơ sở căn cứ vào một (hay một số) tiêu thức nhất định.

b – Ý nghĩa của phân tổ thống kê

- Được dùng nhiều trong các cuộc điều tra thống kê, đặc biệt là điều tra không toàn bộ.
- Là phương pháp cơ bản để tiến hành tổng hợp thống kê.
- Là một trong các phương pháp quan trọng của phân tích thống kê.

c - Nhiệm vụ của phân tổ thống kê

- Phân chia hiện tượng nghiên cứu thành các loại hình khác nhau.
- Nghiên cứu kết cấu của hiện tượng
- Nghiên cứu mối liên hệ giữa các tiêu thức.

2 – Tiêu thức phân tổ

a – KN :

Là tiêu thức được chọn làm căn cứ để phân tổ TK.

b – Các nguyên tắc lựa chọn tiêu thức phân tổ

- Căn cứ vào mục tiêu nghiên cứu
- Căn cứ vào tính chất, đặc điểm của đối tượng nghiên cứu
- Căn cứ vào thời gian nghiên cứu
- Căn cứ vào khả năng của đơn vị.

3 – Xác định số tổ

a – TH1: *Tiêu thức phân tổ có ít biểu hiện hoặc lượng biến của tiêu thức thay đổi ít.*

Cách xác định số tổ :

Coi mỗi biểu hiện hoặc mỗi lượng biến là cơ sở hình thành một tổ.

VD : Phân tổ SV theo giới tính

Phân tổ CN theo số máy 1 CN phụ trách ...

3 – Xác định số tổ

b – TH2 : *Tiêu thức phân tổ có nhiều biểu hiện hoặc lượng biến của tiêu thức thay đổi lớn.*

- Đối với tiêu thức có nhiều biểu hiện :
Tiến hành ghép những biểu hiện tương tự nhau thành một tổ.

- Đối với tiêu thức số lượng có lượng biến thay đổi lớn :

Dựa trên QH lượng chất để phân tổ. (Lượng biến đổi đến mức độ nào thì làm chất biến đổi, mỗi khi chất thay đổi hình thành 1 tổ).

VD : Điểm học tập của sinh viên chia thành :

9 – 10 : Xuất sắc

8 – 9 : Giỏi

7 – 8 : Khá

5 – 7 : TB

3 – 5 : Yếu

< 3 : Kém

Trong những TH này, mỗi tổ sẽ gồm 1 phạm vi lượng biến có 2 giới hạn rõ rệt.

+ Lượng biến nhỏ nhất của tổ làm cho tổ đó hình thành ($x_{i \min}$) gọi là giới hạn dưới của tổ.

+ Lượng biến lớn nhất của tổ mà vượt qua giới hạn đó sẽ chuyển sang tổ khác ($x_{i \max}$) gọi là giới hạn trên của tổ.

Chênh lệch giữa giới hạn trên và giới hạn dưới của tổ gọi là khoảng cách tổ (h_i).

$$h_i = x_{i \max} - x_{i \min}$$

Phân tổ có giới hạn gọi là phân tổ có khoảng cách tổ.

- Nếu khoảng cách tổ bằng nhau có thể tính khoảng cách tổ bằng CT :

$$h = (X_{\max} - X_{\min}) : n$$

h : trị số k/c tổ

X_{\max} , X_{\min} : Lượng biến lớn nhất và lượng biến nhỏ nhất trong tổng thể.

n : Số tổ

Phân tổ với khoảng cách tổ bằng nhau thường dùng khi lượng biến thay đổi một cách đều đặn.

VD1 : Nếu chia TN thành 4 tổ với khoảng cách tổ bằng nhau :

$$h = (3000 - 2200) : 4 = 200 \text{ (1000đ)}$$

Hình thành các tổ (class):

2200 – 2400

2400 – 2600

2600 – 2800

2800 – 3000

Khi chia tổ theo CT trên, giới hạn trên của tổ đứng trước bằng giới hạn dưới của tổ đứng sau.

Chú ý :

- Thực tế, khoảng cách tổ thường lấy số tròn nên khi tính h có thể điều chỉnh các trị số của lượng biến (X_{\max} , X_{\min}) trong CT:

VD : $X_{\max} = 45$; $X_{\min} = 2$; $n = 4$. Ta có thể tính

$$h = (46 - 2) : 4 = 11 \text{ hoặc } (45 - 1) : 4 = 11$$

- TH tổ thứ nhất hoặc tổ cuối cùng không có giới hạn dưới hoặc giới hạn trên thì tổ đó gọi là tổ mở.
- + Đối với tiêu thức số lượng : Tổ mở hay được sử dụng trong TH không biết rõ lượng biến lớn nhất hoặc lượng biến nhỏ nhất của tiêu thức phân tổ.
- + Đối với tiêu thức thuộc tính : Tổ mở được sử dụng khi không có đầy đủ thông tin chi tiết về tất cả các tổ hoặc nếu có thì sẽ quá nhiều tổ. TH này tổ mở thường được gọi là “loại khác” và bao gồm tất cả các thông tin chi tiết chưa được liệt kê ở các tổ trên.

4 – Dãy số phân phối

a – KN : Là dãy số được tạo ra khi tiến hành phân chia các đơn vị của 1 hiện tượng KT-XH theo một tiêu thức nào đó.

Các loại dãy số phân phối :

- Dãy số thuộc tính : Tổng thể được phân tổ theo tiêu thức thuộc tính.
- Dãy số lượng biến : Tổng thể được phân tổ theo tiêu thức số lượng.

b- Cấu tạo :

Dãy số phân phối gồm 2 thành phần:

- Các biểu hiện hoặc các lượng biến của tiêu thức phân tổ (kí hiệu : x_i).
- Tần số tương ứng (kí hiệu : f_i).

Tần số là số lần lặp lại của một biểu hiện hoặc một lượng biến nào đó hay chính là số đơn vị của tổng thể được phân phối vào mỗi tổ.

c - Một số khái niệm khác

+ Tần suất (d_i) : Là tần số được biểu hiện bằng số tương đối (% , lần).

Ý nghĩa : Cho biết số đơn vị mỗi tổ chiếm bao nhiêu % trong toàn bộ tổng thể.

$$d_i = \frac{f_i}{\sum f_i}$$

Nếu d_i tính bằng lần : $\sum d_i = 1$

Nếu d_i tính bằng % : $\sum d_i = 100$

+ Tần số tích lũy (Si)

- *Tần số tích lũy tiến là tổng các tần số khi ta cộng dồn từ trên xuống.*

x_i	f_i	d_i	S_i
x_1	f_1	$f_1 / \sum f_i$	f_1
x_2	f_2	$f_2 / \sum f_i$	$f_1 + f_2$
x_3	f_3	$f_3 / \sum f_i$	$f_1 + f_2 + f_3$
\dots	\dots	\dots	\dots
x_n	f_n	$f_n / \sum f_i$	$\sum f_i$

+ Tần số tích lũy (Si)

VD: có 100 suất học bổng cho SV có KQ học tập tốt

Điểm bình quân (điểm)	Tần số (SV)	Tần số tích lũy (SV)
9,0 trở lên	5	5
8,8 – 9,0	15	15
8,5 – 8,8	30	50
8,3 – 8,5	52	102
8,0 – 8,3	150	252
...

- Tác dụng: (Đối với dãy số lượng biến)
 - + **TH không có khoảng cách tổ** : Tần số cho biết số đơn vị của tổng thể có lượng biến nhỏ hơn hoặc bằng lượng biến của tổ đó.
 - + **TH có khoảng cách tổ** : Tần số tích lũy phản ánh số đơn vị tổng thể có lượng biến nhỏ hơn giới hạn trên của tổ đó.

+ VD : Phân tổ các hộ gia đình theo diện tích nhà ở :

DT nhà (m^2)(x_i)	Số hộ (f_i)	S_i
< 10	5	5
$10 - 30$	10	15
$30 - 50$	30	45
$50 - 70$	40	85
≥ 70	15	100

$S_i = 85$ cho biết có 85 hộ gia đình có DT $< 70m^2$

+ Mật độ phân phối (Di)

Mật độ phân phối là tỉ số giữa tần số với trị số khoảng cách tổ.

Công thức:

$$D_i = \frac{f_i}{h_i}$$

VD :

NSLĐ (chiếc)	Số CN	hi	Di
30 – 40	30	10	3
40 – 50	50	10	5
50 – 70	80	20	4
70 – 75	35	5	7

Thời gian cần thiết (tính bằng giây) để hoàn thiện 1 sản phẩm của 40 công nhân thuộc phân xưởng hoàn thiện được theo dõi như sau:

69	65	72	71	60	61	74	73
75	73	65	68	74	72	63	67
68	72	69	66	65	73	73	69
75	76	70	70	63	64	60	67
61	62	69	74	68	73	77	74

Yêu cầu

- Dùng phương pháp biểu đồ cảnh lá để tổng hợp
- Xây dựng bảng phân tổ với số tổ là 6 và khoảng cách tổ đều nhau

KL :

- Các bước cơ bản để tiến hành phân tổ (Phân tổ theo một tiêu thức hay còn gọi là phân tổ giản đơn):
 - + Chọn tiêu thức phân tổ
 - + Xác định số tổ (và khoảng cách tổ)
 - + Sắp xếp các đơn vị vào các tổ tương ứng (XD dãy số phân phối)

- Các bước cơ bản để tiến hành phân tổ theo nhiều tiêu thức (phân tổ kết hợp):
 - + Lựa chọn tiêu thức phân tổ : Liệt kê những tiêu thức phân tổ và sắp xếp các tiêu thức phân tổ đó theo thứ tự hợp lý để dễ phân tích và nhận xét.
 - + Xác định số tổ của mỗi tiêu thức
 - + Chọn các đơn vị vào các tổ và các tiêu tổ tương ứng.

Bài tập

Có tài liệu theo dõi thời gian thực hiện HĐ của một doanh nghiệp xuất khẩu (*đv: ngày*)

8	10	19	20	9	26	16	12
14	16	19	12	17	18	7	14
6	13	3	23	4	16	20	20
7	5	17	11	15	6	10	21
17	4	15	22	27	11	19	18
21	18	9	19	14	21	17	8

Yêu cầu

- Xây dựng bảng phân tổ thời gian thực hiện HĐ với khoảng cách tổ đều nhau bằng 6 ngày
- Nhận xét về thời gian thực hiện HĐ của doanh nghiệp
- Giả sử tại đầu kỳ kinh doanh, doanh nghiệp đặt mục tiêu là thực hiện được 50% số HĐ trong vòng nửa tháng, vậy trong kỳ, doanh nghiệp có thực hiện được mục tiêu này không?

IV - Bảng TK và đồ thị TK

1 - Bảng thống kê

a – KN :

Là bảng trình bày các thông tin TK một cách có hệ thống, hợp lý, rõ ràng nhằm nêu lên những đặc trưng về mặt lượng của hiện tượng nghiên cứu

b- Cấu tạo bảng TK

- Về hình thức : Bảng TK gồm các hàng ngang, cột dọc, các tiêu đề và số liệu

Kết quả sản kinh doanh công ty A giai đoạn 1999-2002

đơn vị: triệu VND

Chỉ tiêu	1999	2000	2001	2002
Doanh thu	12.000	13.500	13.050	13.780
Chi phí	8.400	9.600	9.750	9.860*
Lợi nhuận	3.600	3.900	3.300	3.920

Nguồn: Phòng kế hoạch - Tổng hợp công ty A

** Chưa tính thuế thu nhập đặc biệt*

- Về nội dung : Gồm 2 phần
- + Phần chủ đề (chủ từ) : Trình bày các bộ phận của hiện tượng nghiên cứu...hay có thể là không gian hoặc thời gian nghiên cứu của hiện tượng đó.
- + Phần giải thích (tân từ) : gồm các chỉ tiêu giải thích các đặc điểm của đối tượng nghiên cứu, giải thích cho phần chủ từ.

c- Yêu cầu khi xây dựng bảng TK

- Qui mô bảng không nên quá lớn
- 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.

- 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.

VD:

Giá trị xuất khẩu một số MH của VN tháng 2/2003

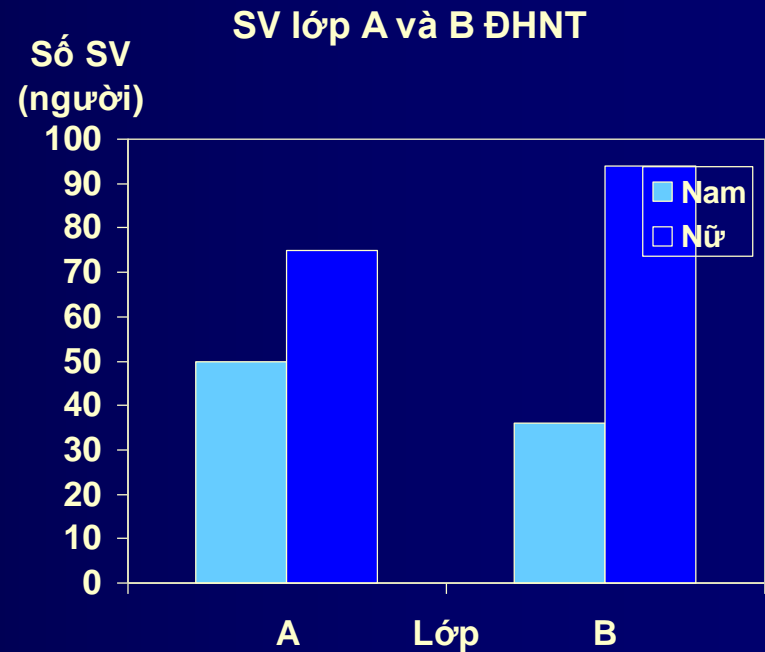
Mặt hàng	Lượng XK (1000 tấn)	Giá trị XK (triệu USD)
1. Gạo
2. Cà phê	50	36
3. Cao su	35	26
4. Dầu thô	1340	292
5. Than đá	300	8

Nguồn: bản tin XNK – BTM số ... tháng 3 năm 2003

2 - Đồ thị thống kê

a - KN :

Là các hình vẽ hoặc đường nét hình học dùng để miêu tả có tính chất qui ước các thông tin thống kê.



b – Tác dụng :

Ứng dụng rộng rãi trong mọi công tác nhằm hình tượng hoá về hiện tượng nghiên cứu, cụ thể biểu hiện:

- + Sự phát triển của hiện tượng qua thời gian
- + Kết cấu và biến động kết cấu của hiện tượng
- + Tình hình thực hiện kế hoạch
- + Mối liên hệ giữa các hiện tượng

.....

c– Các loại đồ thị TK

+ Căn cứ theo nội dung phản ánh:

+ Đồ thị phát triển

+ Đồ thị kết cấu

+ Đồ thị liên hệ

+ Đồ thị so sánh

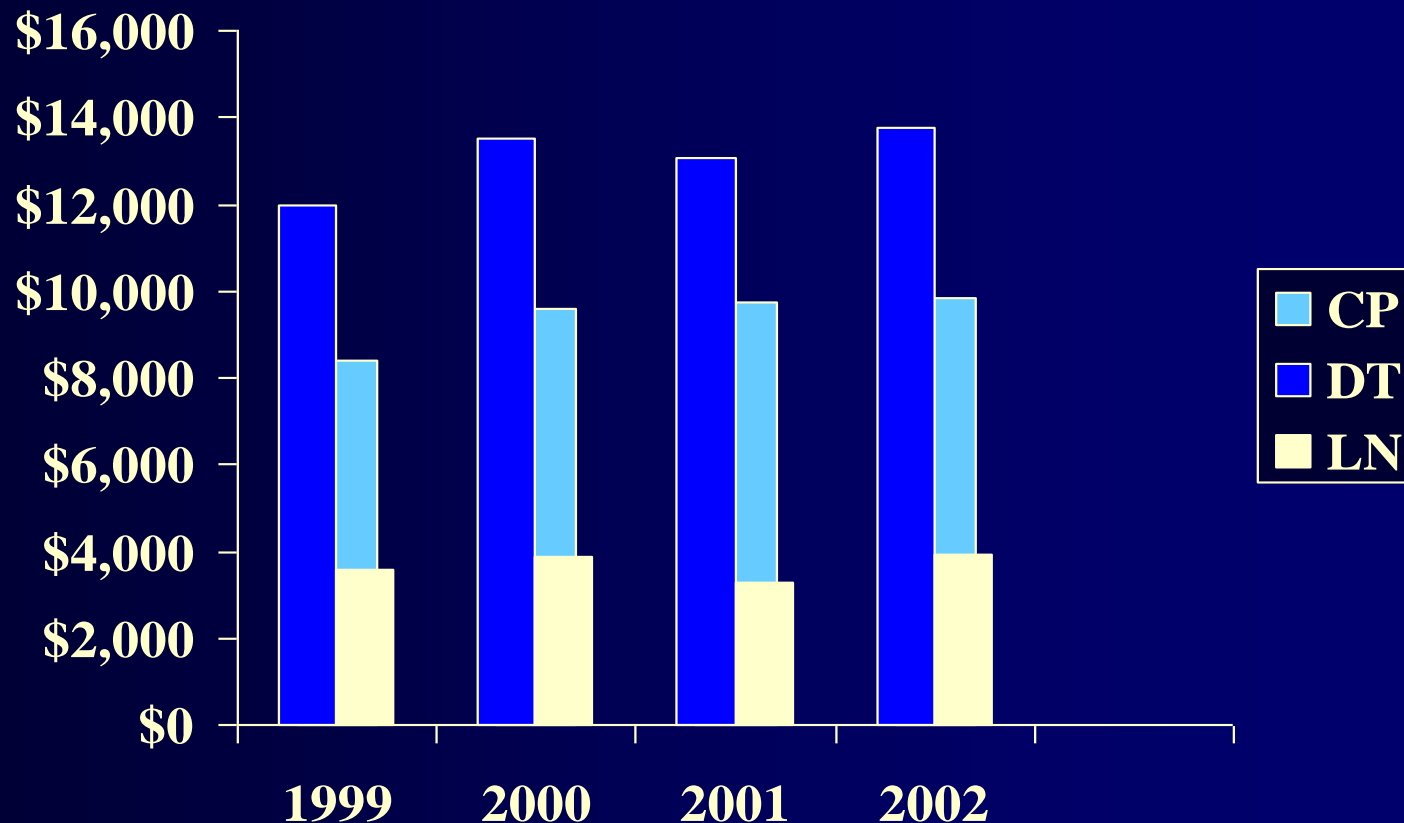
+ Đồ thị phân phối

+ Đồ thị hoàn thành kế hoạch

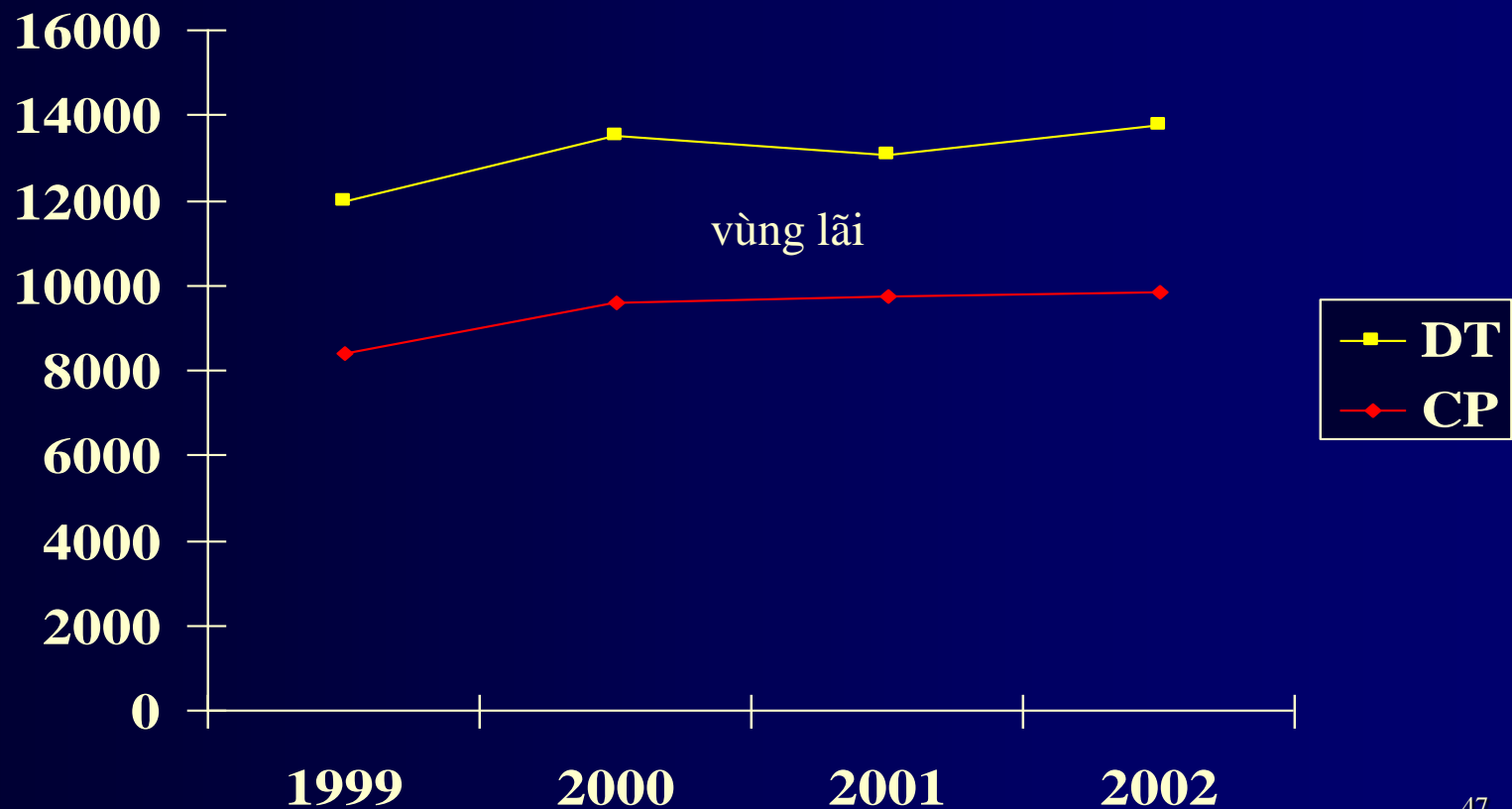
.....

- Căn cứ vào hình thức biểu hiện:
- + Biểu đồ hình cột
- + Biểu đồ tượng hình (biểu hiện bằng các hình vẽ tượng trưng, dùng để tuyên truyền, cổ động...)
- + Biểu đồ diện tích (hình vuông, hình tròn, hình chữ nhật...)
- + Đồ thị đường gấp khúc
- + Bản đồ thống kê

Ví dụ: Biểu đồ kết quả kinh doanh công ty A 1999 - 2002



Biểu đồ kết quả kinh doanh công ty A 1999 - 2002



Có tài liệu về mối liên hệ giữa trình độ và thu nhập của 30 công dân như sau:

TN (\$/year)	Trình độ	TN (\$/year)	Trình độ	TN (\$/year)	Trình độ
17000	H.C	21200	B.S	17200	2ysCollege
20800	B.S	28000	B.S	19600	B.A
27000	M.A	30200	H.C	36200	M.S
70000	M.D	22400	2ysCollege	14400	1ysCollege
29000	Ph.D	100000	M.D	18400	2ysCollege
14400	10 th grade	76000	Law degree	34400	B.A
19000	H.C	44000	Ph.D	26000	H.C
23200	M.A	17600	11 th grade	52000	Law degree
30400	H.C	25800	H.C	64000	Ph.D
25600	B.A	20200	1ysCollege	32800	B.S ⁴⁸