Hai nhóm cá có khối lượng được đo theo cùng một đơn vị đo, khối lượng trung bình của chúng xấp xỉ nhau. Nhóm cá thứ 2 có phương sai bé hơn. Từ đó suy ra rằng nhóm cá thứ 2 có khối lượng đồng đều hơn.

5. $\overline{x} = 34\ 087\ 500$ đồng.

Sắp thứ tự cho các số liệu đã cho, ta thu được dãy không giảm số liệu sau

20060, 20110, 20350, 20350, 20910, 20960, 21130, 21360, 21410, 21410, 76000, 125 000 (nghìn đồng).

Từ đó ta có

$$M_e = \frac{20960 + 21130}{2} = 21045 \text{ (nghìn đồng).}$$

Trong các số liệu thống kê đã cho có sự chênh lệch nhau rất lớn, nên số trung vị ($M_e = 21\ 045\ 000$đ) được chọn làm đại diện cho mức lương hàng năm của mỗi người trong 12 cán bộ và nhân viên của công ti đã được khảo sát.

6. a) Mốt là mẫu 1.

b) Trong sản xuất, nhà máy nên ưu tiên cho mẫu 1.

7. (C).

8. (B).

9. (C).

10. (D).

11. (A).

KIỂM TRA CHƯƠNG V (1 tiết)

Gợi ý để kiểm tra cuối chương V

ĐỀ SỐ 1 *(45 phút)*

Cho bảng phân bố tần số ghép lớp sau (Bảng 12)

Đáp án

Câu 1 (Bảng 13)

Thành tích nhảy xa của 45 học sinh lớp 6A trường Trung học cơ sở T

Lớp thành tích (m)	Tần suất (%)
[2,2 ; 2,4)	6,67
[2,4 ; 2,6)	13,33
[2,6 ; 2,8)	26,67
[2,8 ; 3,0)	24,44
[3,0 ; 3,2)	17,78
[3,2 ; 3,4]	11,11
Cộng	100 (%)

Bảng 13

Câu 2

Trong 45 học sinh được khảo sát, ta thấy

Chiếm tỉ lệ thấp nhất (6,67%) là những học sinh có thành tích nhảy xa từ 2,2 m đến dưới 2,4 m ;

Chiếm tỉ lệ cao nhất (26,67%) là những học sinh có thành tích nhảy xa từ 2,6 m đến dưới 2,8 m ;

Đa số (68,89%) học sinh có thành tích nhảy xa từ 2,6 m đến dưới 3,2 m.

Câu 3

1. $\overline{x} \approx 2{,}8$m ; $s^2 \approx 0{,}1$; $s \approx 0{,}28$ m.

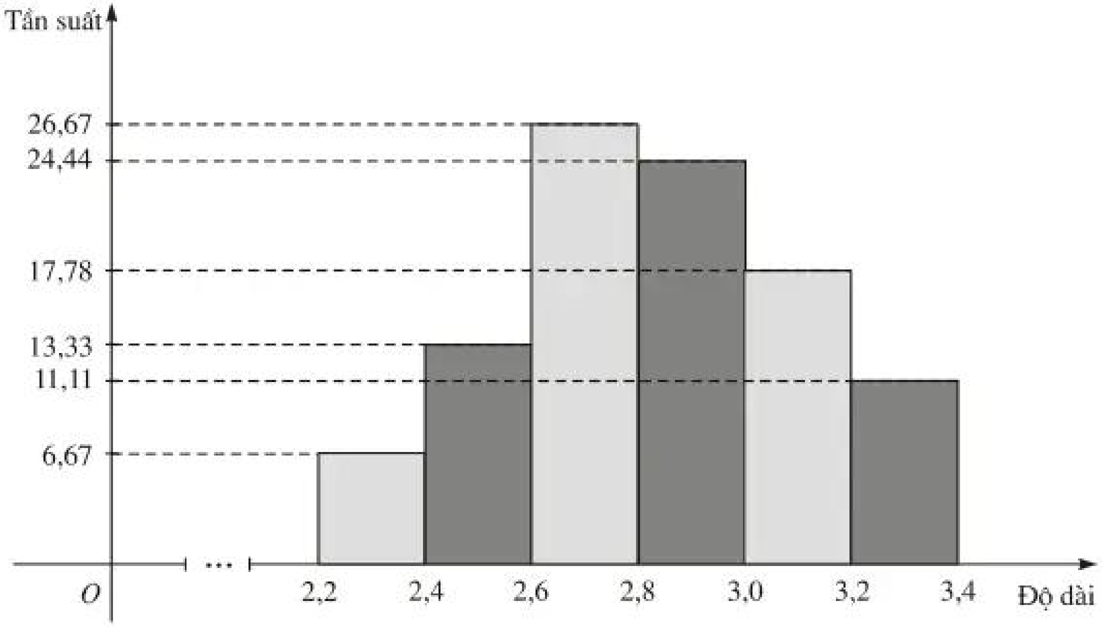
2. Ở lớp 6A : $\overline{x} \approx 2{,}8$ m ; $s^2 \approx 0{,}1$.

   Ở lớp 6B : $\overline{x} \approx 3{,}0$ m ; $s^2 \approx 0{,}3$.

   Ở lớp 6C : $\overline{x} \approx 3{,}0$ m ; $s^2 \approx 0{,}2$.

Do đó, trong trường hợp được xét, thành tích nhảy xa của học sinh ở hai lớp 6B và 6C là cao bằng nhau và cao hơn ở lớp 6A. Đồng thời, thành tích nhảy xa của các học sinh ở lớp 6C là đều hơn ở lớp 6B.

Câu 4 (h.18)



Hình 18. Biểu đồ tần suất hình cột về thành tích nhảy xa (m) của 45 học sinh lớp 6A trường Trung học cơ sở T

ĐỀ SỐ 2 *(45 phút)*

Câu 1 (8,5 điểm)

Cho bảng phân bố tần số ghép lớp sau (bảng 14).

Khối lượng của 85 con lợn (của đàn lợn I) được xuất chuồng (ở trại nuôi lợn N)

Lớp khối lượng (kg)	Tần số
[45 ; 55)	10
[55 ; 65)	20
[65 ; 75)	35
[75 ; 85)	15
[85 ; 95]	5
Cộng	85

Bảng 14

1. (3 điểm). Hãy lập bảng phân bố tần suất ghép lớp, với các lớp như ở bảng 14.

2. (2 điểm). Dựa vào bảng phân bố tần suất ghép lớp đã lập được, hãy xác định xem

a) Số lợn (của đàn lợn I) có khối lượng từ 75 kg trở lên chiếm bao nhiêu phần trăm ?

b) Số lợn (của đàn lợn I) có khối lượng từ 55 kg đến dưới 85 kg chiếm bao nhiêu phần trăm ?

3. (3,5 điểm).

a) Hãy tính số trung bình cộng, phương sai và độ lệch chuẩn của các số liệu thống kê đã cho.

b) Biết rằng sau đó hai tháng, trại N cho xuất chuồng thêm hai đàn lợn, trong đó

Đàn lợn II có khối lượng trung bình bằng 78 kg và phương sai bằng 100.

Đàn lợn III có khối lượng trung bình bằng 78 kg và phương sai bằng 110.

Hãy so sánh khối lượng của lợn trong ba đàn lợn kể trên.

Câu 2 (1,5 điểm)

Cho biểu đồ hình quạt (h.19)

Cơ cấu (phân loại) tập thể học sinh khối 10 trường Trung học phổ thông H, phân theo điểm thi Toán (cuối năm học 2000 – 2001)

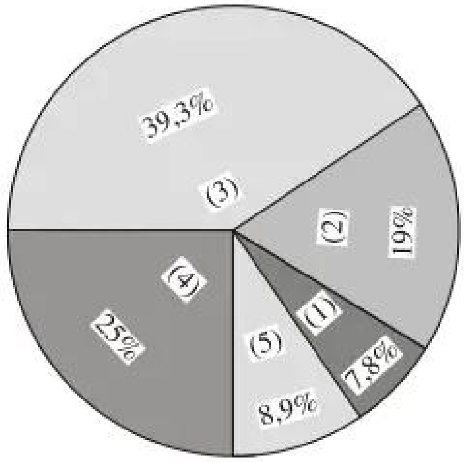
(1) Học sinh kém (đạt từ điểm 1 đến điểm 2).

(2) Học sinh yếu (đạt từ điểm 3 đến điểm 4).

(3) Học sinh trung bình (đạt từ điểm 5 đến điểm 6).

(4) Học sinh khá (đạt từ điểm 7 đến điểm 8).

(5) Học sinh giỏi (đạt từ điểm 9 đến điểm 10).



Hình 19

Dựa vào biểu đồ hình quạt đã cho, hãy lập bảng cơ cấu theo mẫu sau (bảng 15)

Cơ cấu (phân loại) tập thể học sinh khối 10 trường Trung học phổ thông H, phân theo điểm thi Toán (cuối năm học 2000 - 2001)

Các loại học sinh		Số phần trăm
Học sinh kém	(từ 1 điểm đến 2 điểm)	
Học sinh yếu	(từ 3 điểm đến 4 điểm)	
Học sinh trung bình	(từ 5 điểm đến 6 điểm)	
Học sinh khá	(từ 7 điểm đến 8 điểm)	
Học sinh giỏi	(từ 9 điểm đến 10 điểm)	
Cộng		100(%)

Bảng 15

Đáp án

Câu 1

1. *Khối lượng của 85 con lợn (của đàn lợn I) được xuất chuồng (ở trại nuôi lợn N)* (Bảng 16)

Lớp khối lượng (kg)	Tần suất (%)
[45 ; 55)	11,8
[55 ; 65)	23,5
[65 ; 75)	41,2
[75 ; 85)	17,6
[85 ; 95]	5,9
Cộng	100 (%)

Bảng 16

2. a) 23,5% ;   b) 82,3%.

3. a) $\overline{x} \approx 68$ kg ; $s^2 \approx 108$; $s \approx 10,4$ kg.

b) Đàn lợn I có $\overline{x}_1 \approx 68$ kg, $s_1^2 \approx 108$.

Đàn lợn II có $\overline{x}_2 \approx 78$ kg, $s_2^2 = 100$.

Đàn lợn III có $\overline{x}_3 \approx 78$ kg, $s_3^2 = 110$.

Do đó, trong 3 đàn lợn được xét, con lợn ở đàn II và đàn III có *khối lượng bằng nhau* và cao hơn khối lượng của con lợn ở đàn I. Đồng thời khối lượng của các con lợn ở đàn lợn II là đều hơn ở đàn III.

Câu 2 (Bảng 17)

Cơ cấu (phân loại) tập thể học sinh khối 10 trường Trung học phổ thông H, phân theo điểm thi Toán (cuối năm học 2000 - 2001)

Các loại học sinh		Số phần trăm
Học sinh kém	(từ 1 điểm đến 2 điểm)	7,8
Học sinh yếu	(từ 3 điểm đến 4 điểm)	19,0
Học sinh trung bình	(từ 5 điểm đến 6 điểm)	39,3
Học sinh khá	(từ 7 điểm đến 8 điểm)	25,0
Học sinh giỏi	(từ 9 điểm đến 10 điểm)	8,9
Cộng		100(%)

Bảng 17

BỔ SUNG

Ở chương này cần lưu ý đến một số vấn đề sau

I – MỘT SỐ KHÁI NIỆM BAN ĐẦU VỀ THỐNG KÊ Ở LỚP 7

1. Dấu hiệu. Số liệu thống kê

Quá trình nghiên cứu thống kê một hiện tượng kinh tế – xã hội thường được bắt đầu từ "Điều tra thống kê" về một dấu hiệu nào đó, chẳng hạn như

a) Điều tra về "Số đồng hồ treo tường bán được trong mỗi ngày" của tháng giáp tết ở một cửa hàng bách hoá. Ở đây, dấu hiệu điều tra là "Số đồng hồ treo tường bán được trong mỗi ngày".

b) Điều tra về "Khối lượng của mỗi trái xoài" trong 100 trái xoài ở nông trường **N**. Ở đây, dấu hiệu điều tra là : "Khối lượng của mỗi trái xoài".

Các số liệu thu thập được khi điều tra về một dấu hiệu được gọi là các số liệu thống kê. Mỗi số liệu thống kê là một giá trị điều tra được của dấu hiệu (các giá trị này không nhất thiết khác nhau).

Có thể trình bày các số liệu thống kê thành *dãy các số liệu thống kê* hoặc thành *bảng các số liệu thống kê*.

Ví dụ 1

Khi điều tra "Khối lượng chè trong mỗi hộp chè" của 30 hộp chè lấy từ kho của một nhà máy chế biến chè, người ta thu được 30 số liệu ghi trong bảng dưới đây (Bảng 18).

Khối lượng chè của 30 hộp chè (đơn vị đo : g)

101	100	98	98	99	100	99	102	100	100	100	101	99	102	99
101	100	99	100	99	101	101	99	98	102	101	100	100	99	100

Bảng 18

Trong ví dụ này, ta có : Dấu hiệu điều tra (kí hiệu là *X*) là "Khối lượng chè trong mỗi hộp chè" ; mỗi hộp chè là một đơn vị điều tra ; mỗi số liệu thu được khi điều tra một hộp chè là một số liệu thống kê. Tất cả có 30 số liệu thống kê.

2. Tần số. Bảng phân bố tần số

a) Tần số

Trong 30 số liệu thống kê ở bảng 18, ta thấy

Có 5 giá trị khác nhau là (đơn vị : g) : $x_1 = 98$; $x_2 = 99$; $x_3 = 100$; $x_4 = 101$; $x_5 = 102$.

Giá trị $x_1 = 98$ xuất hiện 3 lần. Khi đó ta gọi $n_1 = 3$ là *tần số* của giá trị x_1.

Tương tự, $n_2 = 8$; $n_3 = 10$; $n_4 = 6$; $n_5 = 3$ lần lượt là tần số của các giá trị x_2, x_3, x_4, x_5.

Tần số của một giá trị trong các giá trị khác nhau của một dãy các số liệu thống kê là số lần xuất hiện giá trị đó trong dãy.

b) Bảng phân bố tần số

Dựa vào các kết quả thu được ở trên, ta lập bảng sau (Bảng 19)

Khối lượng chè của 30 hộp chè

Khối lượng chè (g)	98	99	100	101	102	Cộng
Tần số	3	8	10	6	3	30

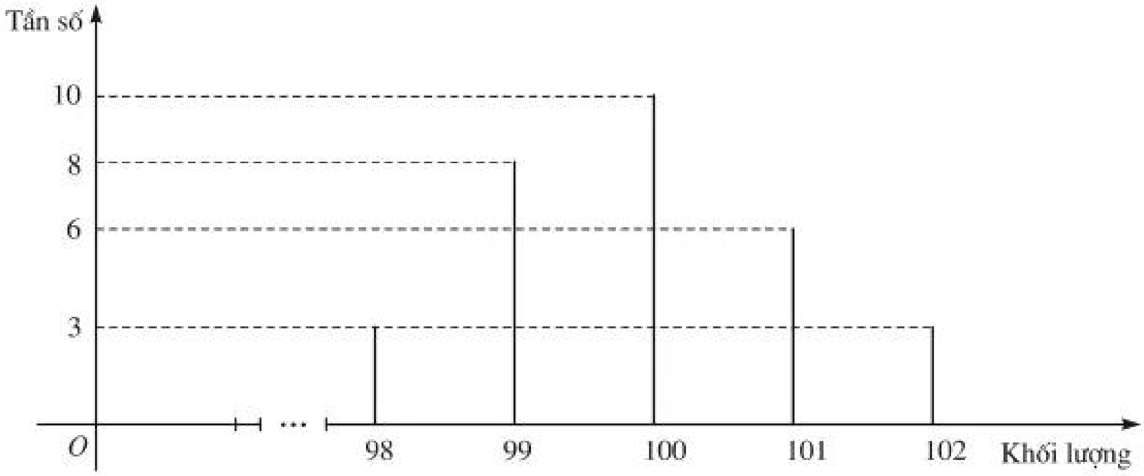
Bảng 19

Bảng 19 phản ánh tình hình phân bố của các số liệu thống kê đã cho được gọi là *bảng phân bố tần số*.

3. Biểu đồ đoạn thẳng tần số

Ví dụ 2

Xét bảng 19, để có được hình ảnh trực quan về tình hình phân bố của các số liệu thống kê, người ta mô tả bảng 19 bằng cách vẽ biểu đồ đoạn thẳng tần số sau đây (h.20)



Hình 20. Biểu đồ đoạn thẳng tần số về khối lượng (g) chè của 30 hộp chè

(Ở hình 20, trên trục Ox, phần "•••" biểu diễn phần bị cắt bỏ của hình vẽ của đoạn thẳng [0 ; 98], để rút ngắn hình của đoạn này).

4. Số trung bình cộng

Dựa vào bảng 19, trung bình cộng của các số liệu thống kê được tính như sau

$$\overline{x} = \frac{3 \times 98 + 8 \times 99 + 10 \times 100 + 6 \times 101 + 3 \times 102}{30} \approx 99{,}93\ (\text{g}).$$

Với kết quả này, có thể lấy giá trị $\overline{x} = 99{,}93$ g làm đại diện cho khối lượng chè của mỗi hộp chè đã được khảo sát.

5. Mốt

Trong bảng 19, giá trị $x = 100$ là *giá trị có tần số cao nhất* (tần số đó bằng 10). Người ta gọi giá trị 100 là mốt của bảng 19 (còn gọi là mốt của dấu hiệu X), kí hiệu là M_O và viết $M_O = 100$.

Từ kết quả vừa thu được, ta nói

> Trong 30 hộp chè được khảo sát, gặp nhiều nhất là những hộp chè có khối lượng chè bằng 100 gam.

Thông tin này là rất có ích cho công tác quản lí sản xuất của nhà máy.

II – BA LOẠI BIẾN

1. Biến định lượng

Biến định lượng là biến mà ta có thể xác định được các giá trị của nó (gọi là các giá trị định lượng) nhờ phép đo theo đơn vị có sẵn hoặc quy ước. Chẳng hạn, chiều cao của mỗi người (trong một nhóm người nào đó) là một biến định lượng.

Biến định lượng X được gọi là rời rạc nếu X chỉ có thể nhận giá trị trong một tập hợp hữu hạn hoặc trong một tập hợp vô hạn đếm được. Chẳng hạn như nếu X là số lượng sách của mỗi cửa hàng bán sách ở phố A thì X là biến định lượng rời rạc.

Biến định lượng X được gọi là liên tục, nếu X có thể nhận giá trị bất kì trong một khoảng mở rộng nào đó. Chẳng hạn, nếu X là chiều dài của trục giữa xe đạp được tiện ra bởi máy M, thì X là biến định lượng và có thể nhận giá trị bất kì trong khoảng $(d - \alpha\ ;\ d + \alpha)$ trong đó d là chiều dài quy định, α là sai số cho phép. Vì vậy, X là biến định lượng liên tục.

Với biến định lượng liên tục, áp dụng được tất cả các phép toán cộng, trừ, nhân, chia, áp dụng được tất cả các phương pháp thống kê hiện có.

2. Biến định hạng

Biến định hạng là biến mà ta có thể gán cho nó các giá trị (gọi là giá trị định hạng), để xếp đặt các kết quả điều tra được theo *từng hạng có thứ tự*. Chẳng hạn, kết quả học tập của mỗi học sinh (kém ; yếu ; trung bình ; khá ; giỏi) là một biến định hạng.

Với biến định hạng, áp dụng được phép so sánh thứ tự và các phương pháp thống kê chỉ dựa trên quan hệ thứ tự.

3. Biến định tính

Biến định tính là biến mà ta có thể gán cho nó các giá trị (gọi là giá trị định tính), để phân loại các kết quả điều tra được. Chẳng hạn, tình trạng hôn nhân (độc thân ; có gia đình ; li hôn ; goá vợ hoặc goá chồng) là một biến định tính.

Với biến định tính, phương pháp toán học có thể áp dụng được là phép đếm.

4. Chú ý

Mỗi biến định hạng đều quy được về biến định tính. Mỗi biến định lượng đều quy được về biến định hạng. Ngược lại chưa chắc đã đúng.

Điểm số không phải là biến định lượng. Nhưng vì nhiều lí do, trong khoa học Giáo dục người ta quy ước coi điểm số là biến định lượng.

III – VỀ CÁC PHƯƠNG PHÁP TRÌNH BÀY CÁC SỐ LIỆU THỐNG KÊ

1. Phương pháp sử dụng bảng thống kê

Có nhiều loại bảng thống kê, trong đó thông dụng nhất là các bảng phân bố tần số, tần suất ghép lớp. Sau đây là một số vấn đề về bảng phân bố tần suất ghép lớp.

a) Những trường hợp nên lập bảng phân bố ghép lớp

Trường hợp 1. Khi dấu hiệu được điều tra nghiên cứu là biến định lượng liên tục thì nên lập bảng phân bố ghép lớp.

Trường hợp 2. Nếu dấu hiệu được điều tra nghiên cứu là biến định lượng rời rạc, nhưng trong các số liệu thống kê có nhiều số liệu khác nhau làm cho bảng phân bố rời rạc cồng kềnh, khó sử dụng, thì nên lập bảng phân bố ghép lớp.

Trường hợp 3. Tập hợp các phần tử được đưa vào điều tra thống kê bao gồm nhiều loại hình khác nhau (xét theo dấu hiệu được nghiên cứu), mà ta muốn làm nổi bật những đặc điểm riêng của các loại hình đó và những mối liên hệ có trong chúng. Khi đó, có thể sử dụng bảng phân bố ghép lớp để trình bày và phân loại các số liệu thống kê.

Ví dụ. Nếu các số liệu thống kê là điểm số biểu thị kết quả học tập của học sinh trong một lớp học, thì để trình bày các số liệu thống kê, có thể lập bảng phân bố ghép lớp theo 5 lớp sau

Kém (1 điểm ; 2 điểm) ;
Yếu (3 điểm ; 4 điểm) ;
Trung bình (5 điểm ; 6 điểm) ;
Khá (7 điểm ; 8 điểm) ;
Giỏi (9 điểm ; 10 điểm).

b) Cách phân lớp

Trường hợp biến định tính (kể cả khi là biến định hạng)

Trong trường hợp này người ta ghép những loại hình giống nhau hoặc gần giống nhau vào cùng một lớp sao cho không có hai lớp nào giao nhau.

Trường hợp biến định lượng

Trong trường hợp này ta phân lớp theo các bước như sau

Bước 1. Xác định đoạn chứa các số liệu thống kê

Tìm số lớn nhất trong các số liệu thống kê đã cho (kí hiệu là $x_{\max}$) ; tìm số nhỏ nhất trong các số liệu thống kê đã cho (kí hiệu là $x_{\min}$). Như vậy các số liệu thống kê đã cho là nằm trong đoạn $[a\ ;\ b]$, với $a \le x_{\min}$, $b \ge x_{\max}$ và các số a, b được chọn sao cho trong các lớp được phân ra không có lớp nào rỗng (không chứa số liệu thống kê nào).

Bước 2. Xác định số lượng các lớp được phân (kí hiệu là k), và phân đoạn $[a\ ;\ b]$ thành k lớp không giao nhau (là $k - 1$ nửa khoảng và một đoạn) sau đây

$$[x_1;x_2),\ [x_2;x_3),\ldots,\ [x_{k-1};x_k),\ [x_k;x_{k+1}],\ \text{ với }\ x_1 = a,\ x_{k+1} = b.$$

Mỗi một số liệu thống kê đã cho thuộc vào một và chỉ một lớp trong k lớp kể trên.

➢ **Chú ý**

Nếu dấu hiệu được nghiên cứu là biến định lượng liên tục và nếu phân thành k lớp có bề rộng bằng nhau, thì bề rộng h của mỗi lớp được xác định theo công thức

$$h = \frac{b-a}{k}.$$

Bề rộng của một lớp bất kì là hiệu số của hai đầu mút của nó.

*Nếu dấu hiệu được nghiên cứu là biến định lượng rời rạc thì bề rộng h của một lớp là **số các số liệu thống kê khác nhau nằm trong lớp đó**.*

c) Kinh nghiệm trong lập bảng phân bố ghép lớp

Với biến định lượng có thể phân lớp với các lớp có bề rộng bằng nhau hoặc không bằng nhau.

Cần phân lớp sao cho mỗi lớp đều bao hàm những số liệu được xác định trên những phần tử có cùng tính chất (hoặc có tính chất gần giống nhau), và sao cho không có lớp nào có tần số (hoặc tần suất) bằng 0.

Rõ ràng là bảng phân bố ghép lớp đã làm mất đi một số thông tin chi tiết (đó là một số thông tin về từng số liệu riêng lẻ), nhưng đã làm nổi bật được một số thông tin tổng quát hơn, chủ yếu hơn (đó là những thông tin về các lớp được phân ra). Từ đó, có câu hỏi : Phải phân thành bao nhiêu lớp và phân thành những lớp nào, để bảng phân bố vừa gọn, vừa không làm mất đi nhiều thông tin, vừa làm rõ được những thông tin chủ yếu, cần thiết, đồng thời lại *giảm nhẹ được việc tính toán ?* Muốn trả lời câu hỏi này, phải dựa vào tình huống cụ thể của thực tiễn. Tuy nhiên, có thể phân lớp theo những gợi ý sau đây

Số lượng các phần tử được đưa vào điều tra	Số lượng các lớp nên phân ra
Từ 25 đến 60	Từ 4 đến 8 lớp
Từ 60 đến 100	Từ 7 đến 10 lớp
Lớn hơn hay bằng 100	Từ 8 đến 12 lớp

2. Phương pháp đồ thị

Đồ thị thống kê là các hình vẽ hoặc đường nét hình học, dùng để mô tả có tính chất quy ước các tài liệu thống kê khác một cách trực quan có hình ảnh.

Để có được những hình ảnh trực quan về tình hình phân bố của các số liệu thống kê, người ta mô tả các bảng phân bố bằng các đồ thị thống kê gồm các biểu đồ, các đường gấp khúc và những đồ thị thống kê khác.

Việc mô tả các bảng phân bố bằng các đồ thị thống kê còn được gọi là "Trình bày các số liệu thống kê bằng các đồ thị thống kê".

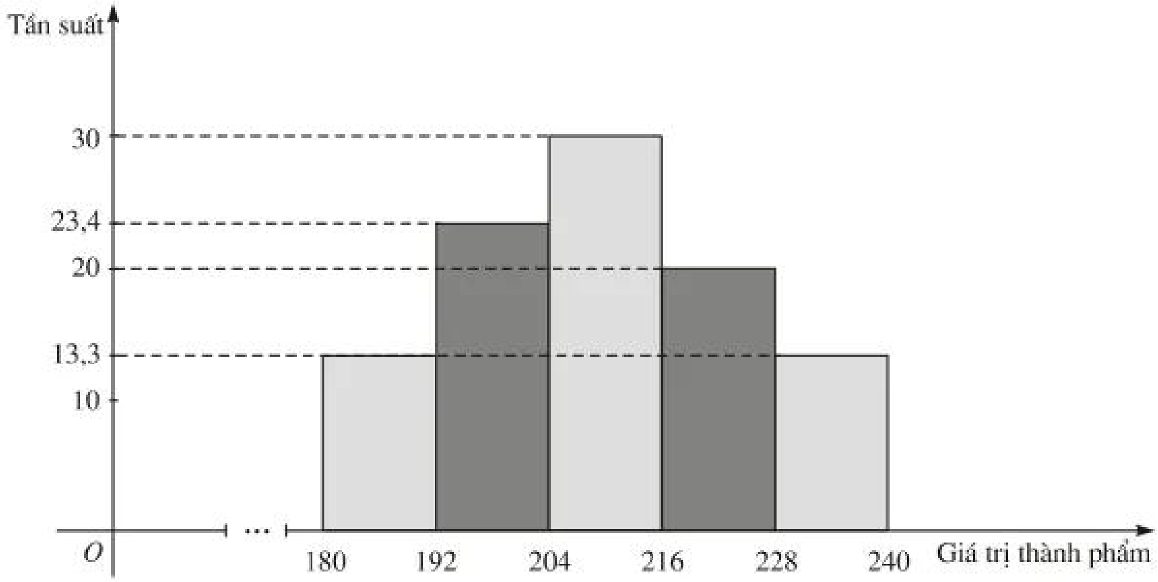
Sau đây là một số vấn đề về cách vẽ "Biểu đồ tần suất hình cột" và "Biểu đồ hình quạt".

a) Cách vẽ biểu đồ tần suất hình cột

Để mô tả bảng phân bố tần suất ghép lớp, ta vẽ biểu đồ tần suất hình cột như sau

Chọn hệ trục toạ độ vuông góc Oxf với đơn vị trên trục hoành Ox là đơn vị của dấu hiệu X được nghiên cứu, đơn vị trên trục tung Of là 1%.

Để đồ thị thống kê cân đối, đôi khi phải cắt bỏ một đoạn nào đó của trục hoành (hoặc của trục tung), chẳng hạn như hình 21, trên trục hoành để rút ngắn hình vẽ của đoạn [0 ; 180], người ta cắt bỏ một phần hình vẽ của đoạn này ; phần "•••" là biểu diễn cho phần hình vẽ bị cắt bỏ.



Hình 21. Biểu đồ tần suất hình cột về giá trị thành phẩm trong mỗi ngày sản xuất, quy ra tiền (nghìn đồng) của 30 ngày được khảo sát ở phân xưởng A

Trên trục hoành đặt các khoảng có các mút biểu diễn cho các mút của các lớp ở bảng phân bố tần suất (độ dài của các khoảng bằng bề rộng của các lớp). Ta gọi các khoảng và các lớp này là tương ứng với nhau.

Lấy các khoảng nói trên làm các cạnh đáy, vẽ các hình chữ nhật có độ dài của các đường cao bằng tần suất của các lớp tương ứng, và nằm về phía chiều dương của trục tung (h.21).

Người ta gọi các hình chữ nhật vừa vẽ được là *biểu đồ tần suất hình cột,* để mô tả bảng phân bố tần suất ghép lớp và trình bày các số liệu thống kê.

b) Cách vẽ biểu đồ hình quạt

Có thể vẽ biểu đồ hình quạt (để mô tả bảng cơ cấu gồm k nhóm) theo cách sau đây

Vẽ một đường tròn.

Coi toàn bộ hình tròn (ứng với cung 360°) là biểu diễn cho $100\% = 1$, ta có hình quạt ứng với cung có số đo độ bằng α sẽ biểu diễn cho số phần trăm

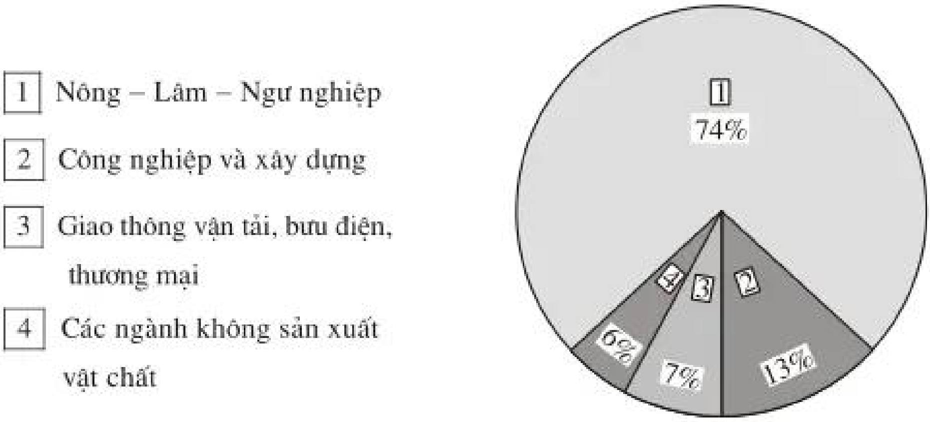
$f = \dfrac{\alpha}{360^{\circ}}$. Do đó, mỗi nhóm thứ i của bảng cơ cấu ứng với số phần trăm

$f_i \left(\sum_{i=1}^{k} f_i = 100\% \right)$ sẽ được biểu diễn bởi một hình quạt mà cung tròn tương

ứng có số đo độ bằng

$$\boxed{\alpha_i = f_i.360^{\circ}} \qquad (1)$$

Vẽ các hình quạt có số đo (độ) của cung tròn tương ứng lần lượt là α_1, α_2, ..., α_i, ..., α_k, với α_i được xác định theo công thức (1). Khi đó, ta thu được một biểu đồ hình quạt mô tả bảng cơ cấu đã cho. Biểu đồ hình quạt ở hình 22 mô tả bảng cơ cấu sử dụng lao động trong các ngành kinh tế của Việt Nam năm 1993.



Hình 22. Biểu đồ hình quạt về cơ cấu sử dụng lao động trong các ngành kinh tế của Việt Nam, năm 1993.

➢ **Chú ý.** *Từ cách vẽ vừa nêu, suy ra rằng số đo độ (và độ dài) của các cung tròn ứng với các hình quạt của biểu đồ tỉ lệ với số phần trăm của các nhóm của bảng cơ cấu.*

IV – VỀ PHƯƠNG PHÁP THU GỌN CÁC SỐ LIỆU THỐNG KÊ

Để thu gọn, đúc kết các số liệu thống kê, người ta thường sử dụng các số đặc trưng của bảng phân bố tần số, tần suất. Các số này phản ánh quy mô và cấu trúc của các số liệu thống kê.

Một bảng phân bố tần số, tần suất có thể có nhiều số đặc trưng khác nhau. Vì lí do sư phạm, SGK chỉ trình bày các số đặc trưng sau đây

Số trung bình cộng, số trung vị, mốt ; phương sai và độ lệch chuẩn.

Sau đây xin bổ sung một số vấn đề có liên quan tới các số đặc trưng nói trên.

1. Về các số định tâm

Số trung bình cộng, số trung vị, mốt biểu thị các cách để xác định vị trí của *điểm trung tâm* của dãy các số liệu thống kê. Chúng được gọi là các số định tâm. Các số này phản ánh xu hướng tập trung của các số liệu thống kê. Bởi vậy, trong những điều kiện nhất định, có thể chọn các số này làm đại diện cho các số liệu thống kê về quy mô và độ lớn.

1) Số trung bình cộng

a) Công thức tính

Trường hợp bảng phân bố tần số, tần suất

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i,$$

trong đó : n_i, f_i là tần số, tần suất của giá trị x_i, n là số các số liệu thống kê $\left(\sum_{i=1}^{k} n_i = n\right)$.

Trường hợp bảng phân bố tần số, tần suất ghép lớp

Trong tính toán, người ta lấy giá trị đại diện của một lớp để thay thế cho mỗi số liệu thống kê thuộc lớp đó. Vì vậy, trong trường hợp này số trung bình cộng được tính (gần đúng) như sau

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \sum_{i=1}^{k} f_i c_i,$$

trong đó : c_i, n_i, f_i lần lượt là giá trị đại diện, tần số, tần suất của lớp thứ i, n là số các số liệu thống kê $\left(\sum_{i=1}^{k} n_i = n\right)$.

b) Về tính đại diện của số trung bình cộng

Nếu các số liệu thống kê là cùng loại và có số lượng n đủ lớn ($n \geq 30$) thì có thể chọn số trung bình cộng làm đại diện cho các số liệu thống kê về quy mô và độ lớn. Tính đại diện của số trung bình cộng được cắt nghĩa một cách trực quan như sau

Ở trung bình cộng của đám đông các số liệu thống kê cùng loại, những tính cách cá biệt của các số liệu thống kê riêng lẻ (được biểu thị ở chỗ hơn kém nhau giữa chúng và ở tính ngẫu nhiên của chúng) sẽ loại trừ nhau và dẫn đến giá trị trung bình cộng $\overline{x}$ có tính tất yếu, biểu thị cho cái chung, cái bản chất nằm trong mỗi số liệu thống kê. Do đó có thể lấy $\overline{x}$ làm đại diện (về quy mô và độ lớn) cho mỗi số liệu trong đám đông các số liệu thống kê cùng loại đã cho.

➢ **Chú ý**

Trong những trường hợp sau đây, không nên chọn số trung bình cộng làm đại diện cho các số liệu thống kê

Các số liệu thống kê không cùng loại. Trường hợp các số liệu thống kê có sự chênh lệch rất lớn là một trường hợp riêng của trường hợp này.

Số lượng các số liệu thống kê quá ít ($n \le 10$).

Dấu hiệu được nghiên cứu không phải biến định lượng.

Và nhiều trường hợp khác nữa.

2) Số trung vị

a) Định nghĩa

Số trung vị của một dãy không giảm (hoặc không tăng) gồm n số liệu thống kê là

Giá trị của số liệu ở chính giữa dãy (số liệu đứng thứ $\frac{n+1}{2}$) nếu n lẻ.

Trung bình cộng của hai số liệu đứng giữa dãy $\left(\text{số liệu đứng thứ } \frac{n}{2} \text{ và số liệu đứng thứ } \frac{n}{2}+1\right)$ nếu n chẵn.

Số trung vị của một dãy số liệu thống kê đã cho là số trung vị của dãy không giảm (hoặc không tăng) được tạo thành từ dãy đã cho.

Trong thực tiễn, với n chẵn, ta thường chọn số trung vị là giá trị của số liệu đứng thứ $\frac{n}{2}$ trong dãy không giảm (hoặc không tăng) được tạo thành từ n số liệu đã cho.

➢ **Chú ý**. *Trong bảng phân bố tần số các số liệu thống kê đã được sắp thứ tự thành dãy không giảm theo các giá trị của chúng. Do đó dễ dàng vận dụng định nghĩa nêu trên để tìm số trung vị của dãy số liệu cho ở bảng phân bố tần số.*

b) Tìm gần đúng số trung vị trong trường hợp bảng phân bố tần số ghép lớp

Trong trường hợp này, ta tìm gần đúng số trung vị theo các bước sau

Bước 1. Tìm lớp chứa trung vị

Tính số $\frac{n}{2}$ (n là số các số liệu thống kê).

Cộng các tần số của các lớp, kể từ lớp thứ 1 đến lớp nào mà tổng tần số bắt đầu vượt qua $\frac{n}{2}$ thì dừng lại.

Lớp dừng lại là lớp chứa trung vị.

Bước 2. Xác định gần đúng số trung vị M_e theo công thức

$$M_e \approx x_t + \frac{h_t}{n_t}\left(\frac{n}{2} - S_t\right)$$

trong đó

x_t là mút trái của lớp chứa trung vị ;

h_t là bề rộng của lớp chứa trung vị ;

n_t là tần số của lớp chứa trung vị.

S_t là tổng các tần số của tất cả những lớp trước khi tới lớp chứa trung vị.

3) Tìm gần đúng mốt trong trường hợp bảng phân bố tần số ghép lớp

Trong trường hợp này, ta tìm gần đúng mốt theo các bước sau

Bước 1. Tìm lớp chứa mốt

Nếu các lớp có bề rộng bằng nhau thì lớp có tần số lớn nhất là lớp chứa mốt.

Nếu các lớp có bề rộng không bằng nhau thì lớp có mật độ phân bố lớn nhất là lớp chứa mốt, trong đó mật độ phân bố của lớp thứ i được kí hiệu là d_i và $d_i = \frac{n_i}{h_i}$, với n_i là tần số của lớp thứ i, h_i là bề rộng của lớp thứ i.

Bước 2. Xác định mốt M_O theo công thức sau

$$M_O = x_O + h_O \frac{\delta_1}{\delta_1 + \delta_2}$$

trong đó

x_O là mút trái của lớp chứa mốt ;

h_O là bề rộng của lớp chứa mốt.

$\delta_1 = n_O - n_{O-1}$ (nếu các lớp có bề rộng bằng nhau) hoặc $\delta_1 = d_O - d_{O-1}$, (nếu các lớp có bề rộng không bằng nhau), với n_O và n_{O-1} là tần số của lớp chứa mốt và lớp liền trước lớp chứa mốt, d_O và d_{O-1} là mật độ phân bố của lớp chứa mốt và lớp liền trước lớp chứa mốt.

$\delta_2 = n_O - n_{O+1}$ (nếu các lớp có bề rộng bằng nhau) hoặc $\delta_2 = d_O - d_{O+1}$ (nếu các lớp có bề rộng không bằng nhau) với n_{O+1} là tần số của lớp liền sau lớp chứa mốt, d_{O+1} là mật độ phân bố của lớp liền sau lớp chứa mốt.

Đôi khi, người ta lấy giá trị đại diện của lớp chứa mốt làm giá trị gần đúng của mốt.

2. Về phương sai và độ lệch chuẩn

Như đã biết, các số định tâm ($\overline{x}$, M_e, M_O) đại diện cho các số liệu thống kê về quy mô và độ lớn. Tuy nhiên, chúng chưa phản ánh được đầy đủ cấu tạo nội bộ của dãy các số liệu thống kê. Ngoài các số định tâm, còn có những số đặc trưng khác được dùng để đánh giá mức độ phân tán của các số liệu thống kê (so với các số định tâm). Trong đó, phương sai và độ lệch chuẩn được dùng để đánh giá mức độ phân tán của các số liệu thống kê (so với số trung bình cộng). Có thể nói rõ thêm về điều này như sau

Để đánh giá mức độ phân tán (so với số trung bình cộng $\overline{x}$) của n số liệu thống kê x_j ; $(j = \overline{1, n})$, trước hết người ta nghĩ đến trung bình cộng của các độ lệch $(x_j - \overline{x})$, nghĩa là nghĩ đến

$$\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x}),$$

với $x_i(i = \overline{1, k})$ là các giá trị khác nhau trong n số liệu thống kê, n_i là tần số của x_i $\left(\text{trong đó } \sum_{i=1}^{k} n_i = n\right)$, nhưng đại lượng này luôn luôn bằng 0. Do đó, người ta nghĩ đến "Độ lệch tuyệt đối trung bình"

$$\frac{1}{n}\sum_{i=1}^{k} n_i\left|x_i - \overline{x}\right|.$$

Sau đó, để tránh dùng giá trị tuyệt đối của các độ lệch, người ta bình phương các độ lệch, rồi lấy trung bình cộng của chúng và thu được "Phương sai"

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2.$$

Nhưng đơn vị đo của s_x^2 là bình phương của đơn vị đo của dấu hiệu X được nghiên cứu. Để tránh điều này, người ta lấy căn bậc hai của s_x^2 và thu được "Độ lệch chuẩn"

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2}.$$

Phương sai và độ lệch chuẩn đều được dùng để đánh giá mức độ phân tán của các số liệu thống kê (so với số trung bình cộng), nhưng khi cần chú ý đến cả đơn vị đo thì phải dùng độ lệch chuẩn.