

# MỘT SỐ VẤN ĐỀ VỀ ÁP DỤNG XÁC SUẤT THỐNG KÊ VÀO THỰC TẾ

• ThS. NGUYỄN PHƯƠNG CHI  
Đại học Sư phạm Hà Nội

Một số kiến thức xác suất, thống kê đã được đưa vào chương trình môn Toán cấp trung học phổ thông (THPT), có nhiều tiềm năng ứng dụng, nhưng tiềm năng này đôi khi còn bị bỏ lỡ chỉ vì việc áp dụng các kiến thức trên liên quan tới những quy trình khá phức tạp. Để khai thác được tiềm năng này, ta nên tránh e dè khi gặp những quy trình như vậy và đôi khi cần cung cấp công cụ hiệu lực để hỗ trợ thực hiện các quy trình đó. Điều này sẽ được trình bày cụ thể đối với các nội dung phân bố tần suất ghép lớp và biểu đồ tần suất hình quạt ở lớp 10 THPT.

## 1. Lập bảng phân bố tần số hoặc tần suất ghép lớp

Theo chuẩn kiến thức, kĩ năng của nội dung này, chương trình chuẩn cũng như chương trình nâng cao của môn Toán cấp THPT đều có ghi chú: "Không yêu cầu biết cách phân lớp; biết đầy đủ các trường hợp lập bảng phân bố tần số - tần suất ghép lớp" ([1], tr. 108 và tr. 175). Việc hạn chế yêu cầu như vậy làm cho học sinh không biết phân lớp một tập dữ liệu, do đó không thể tự mình giải nhiều bài toán thống kê trong đời sống thực tế. Thật vậy, khi số lượng các giá trị khác nhau của dữ liệu điều tra quá lớn làm cho bảng phân bố tần số hoặc tần suất cho từng dữ liệu quá công kềnh, khó sử dụng, ta cần lập bảng phân bố ghép lớp. Ngay cả việc vẽ biểu đồ tần số, tần suất hình cột, biểu đồ tần suất hình quạt và đường gấp khúc tần số, tần suất mà học sinh được học tiếp ngay sau đó cũng đòi hỏi thực hiện bước đầu tiên là lập bảng phân bố tần số, tần suất ghép lớp.

Vi vậy, nên nghiên cứu để trong lần đổi mới chương trình vào thời điểm tới gần đây nhất [2], ghi chú nói trên sẽ được xoá bỏ. Hiện tại với tình cách là một bài đọc thêm đối với học sinh giỏi, hoặc trong chủ đề nâng cao của môn Toán lớp 10 ta có thể hướng dẫn học sinh một cách lập

bảng phân bố tần số, tần suất ghép lớp.

Ta hạn chế việc này ở những biến định lượng (tức là biến mà các giá trị của nó được xác định nhờ phép đo theo đơn vị có sẵn hoặc quy ước) thành một quy trình như sau:

*Bước 1. Quy định một đoạn chứa các số liệu thống kê và phác tính số lớp số liệu*

- Kí hiệu  $\min(x)$  và  $\max(x)$  là giá trị nhỏ nhất và giá trị lớn nhất trong các số liệu thống kê đã cho. Người ta cũng cho phép sử dụng hai giá trị trên với độ xê xích một chút là  $h'$  nên thường đặt  $h' = \frac{1}{2} \min(|x_i - x_j|)$  (là nửa số nhỏ nhất trong các trị tuyệt đối của hiệu số giữa hai số liệu thống kê,  $a' = \min(x) - h'$  và  $b' = \max(x) + h'$ . Ta được một đoạn  $[a'; b']$  chứa toàn bộ các số liệu thống kê.

Nếu các số liệu đều là số tự nhiên thì đặt  $a' = \min(x)$  và  $b' = \max(x)$

- Để chia lớp các số liệu thống kê, trước hết cần quy định số lượng các lớp. Nếu số lớp ít thì hơi nhiều số liệu được gộp vào mỗi lớp làm cho mất mát nhiều thông tin. Ngược lại, nếu số lớp nhiều quá thì bảng số liệu quá dài, rất công kềnh. Tuy số lớp không được quy định bởi một quy tắc chặt chẽ nào, nhưng theo kinh nghiệm thì nên lấy một số phụ thuộc thuận chiều với số lượng  $n$  của toàn bộ các phần tử được đưa vào điều tra. Cụ thể hơn nữa, với  $n$  từ 25 đến 100 phần tử trở lên thì các số liệu thường được chia thành 4 đến 15 lớp, và sẽ rất thuận tiện nếu chia thành 10 lớp. Vậy số lượng  $m'$  của các lớp có thể được phác tính theo kinh nghiệm nói trên căn cứ vào số lượng  $n$  các phần tử được đưa vào điều tra. Số lớp số liệu cũng có thể được quy định dựa vào một thông lệ nào đó trong thực tế, chẳng hạn các điểm số có thể được chia thành 10 lớp hay 5 lớp phù hợp với thang điểm 10 bậc.

*Bước 2. Chia lớp*

Các số liệu được chia lớp theo cách là định ra những nửa khoảng liên tiếp, đôi một không

giao nhau, mỗi khoảng chứa một và chỉ một lớp số liệu được chia ra, cụ thể làm như sau:

- Phác tính bề rộng  $k'$  của nửa khoảng hoặc đoạn chứa mỗi lớp số liệu bằng cách lấy giá trị làm tròn  $k'$  của thương trong phép chia  $b' - a'$  cho  $m'$ .

- Khi thiết kế các đoạn hay nửa khoảng chứa lớp số liệu, người ta thường tránh tạo ra những đoạn hay nửa khoảng chứa lớp số liệu có tần số bằng 0. Muốn vậy ta có thể lấy  $h'$  thay đổi chút ít và lựa chọn  $m'$  để điều chỉnh  $a'$ ,  $b'$ ,  $k'$  cho tới khi thoả mãn yêu cầu trên. Khi đã thoả mãn, ta kí hiệu các giá trị tìm được tương ứng là  $h$ ,  $m$ ,  $a$ ,  $b$  (hai đầu mút của đoạn  $[a, b]$ ) và bề rộng  $k$  của lớp.

- Chia toàn bộ số liệu thống kê (nằm trong đoạn  $[a ; b]$ ) thành  $m$  lớp nằm trong  $m-1$  nửa khoảng và 1 đoạn đôi một không giao nhau được kí hiệu như sau:

$[a ; a+k), [a+k ; a+2k), \dots, [a+(m-2)k ; a+(m-1)k), [a+(m-1)k ; b]$ .

Đó cũng là kí hiệu được dùng để chỉ các lớp số liệu được chia ra.

*Chú ý:* Theo cách trên, gần như toàn bộ số liệu được chia thành những lớp biểu diễn trên các nửa khoảng bằng nhau, mặc dầu về nguyên tắc không bắt buộc phải như vậy, nhất là đối với nửa khoảng đầu tiên và đoạn cuối cùng.

Trường hợp tất cả các số liệu đều là số tự nhiên thì các lớp số liệu có thể được biểu diễn trên các đoạn không giao nhau dưới đây:

$[a ; a+k-1], [a+k ; a+2k-1], \dots, [a+(m-2)k ; a+(m-1)k-1], [a+(m-1)k ; b]$ .

**Bước 3. Lập bảng tần số và tần suất của các lớp**

- Lập bảng có 3 cột với các tiêu đề cột là các lớp, tần số và tần suất.

- Tính tần số và tần suất của mỗi lớp và ghi vào từng ô thích hợp của bảng.

Lớp	[0 ; 2]	[3 ; 5]	[6 ; 8]	[9 ; 11]	[12 ; 14]	[15 ; 17]	[18 ; 19]	Tổng
Tần số	2	6	10	14	8	3	2	45
Tần suất	4,4%	13,3%	22,2%	31,1%	17,8%	6,7%	4,4%	100%

*Ví dụ 1. Thống kê số lỗi trong 1 trang đánh máy vi tính của mỗi người ở một lớp học soạn thảo văn bản gồm 45 người, người ta được bảng sau:*

Số lỗi	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Tần số	0	1	1	3	3	0	2	3	5	4	4	6	3	3	2	2	0	1	1	1

Hãy lập bảng phân bố tần suất ghép lớp về số lỗi trong một trang đánh máy của các học viên lớp học đó.

**Giải:**

**Bước 1. Quy định một đoạn chứa các số liệu thống kê và phác tính số lớp số liệu**

- Giá trị nhỏ nhất và giá trị lớn nhất trong các số liệu là 0 và 19. Tất cả các số liệu trong ví dụ này đều là số tự nhiên. Vậy chọn  $[0 ; 19]$  là một đoạn chứa toàn bộ số liệu thống kê.

- Trong ví dụ này có 45 người được đưa vào điều tra, vậy theo kinh nghiệm đã nêu ở cuối bước 1, số lớp có thể được lấy trong nửa dưới của dãy số từ 4 đến 15, cụ thể là ta có thể chia các số liệu thành 7 lớp.

**Bước 2. Chia lớp**

- Phác tính bề rộng của mỗi nửa khoảng hoặc đoạn chứa mỗi lớp số liệu: kết quả phép chia  $19 : 7$  được làm tròn là 3.

- Căn cứ vào bảng số liệu đã cho, ta thấy ngay cả 6 nửa khoảng và 1 đoạn  $[0 ; 3), [3 ; 6), [6 ; 9), [9 ; 12), [12 ; 15), [15 ; 18), [18 ; 19]$  đều chứa các lớp số liệu có tần số khác 0, vì vậy không cần điều chỉnh số lớp nữa. Ta khẳng định toàn bộ số liệu được chia thành 7 lớp được kí hiệu như trên hoặc dưới dạng sau đây rất thông dụng trong trường hợp tất cả các số liệu đều là số tự nhiên:

$[0 ; 2], [3 ; 5], [6 ; 8], [9 ; 11], [12 ; 14], [15 ; 17], [18 ; 19]$

**Bước 3. Lập bảng tần số và tần suất của các lớp**

Ví dụ 2. Kết quả thi môn Toán học kì 2 năm học 2007- 08 của một lớp 10 THPT được ghi lại vẫn tắt như sau:

Điểm thi	0	1	1,8	2,3	2,5	2,8	3	3,3	3,8	4	4,3	4,5
Tần số	1	1	1	1	2	3	1	3	3	6	3	5
Điểm thi	4,8	5	5,3	5,5	5,8	6	6,3	7	7,3	7,8	9	9,8
Tần số	4	2	2	2	1	2	3	2	2	1	1	1

Hãy lập bảng phân bố tần số và tần suất ghép lớp về điểm thi nói trên.

**Giải:**

**Bước 1. Quy định một đoạn chứa các số liệu thống kê và phác tính số lớp số liệu**

- Giá trị nhỏ nhất và giá trị lớn nhất trong các số liệu là 0 và 9,8. Vì điểm số không thể nhỏ hơn 0 và cũng không thể lớn hơn 10 nên có thể lấy ngay  $[0 ; 10]$  là một đoạn chứa toàn bộ số liệu thống kê.

- Ví dụ này có 53 số liệu, vậy theo kinh nghiệm đã nêu ở cuối bước 1 và chú ý thang điểm 10 bậc, ta có thể chia các số liệu thành 5 lớp.

**Bước 2. Chia lớp**

- Phác tính bề rộng của mỗi nửa khoảng hoặc đoạn chứa mỗi lớp số liệu: kết quả phép chia  $(10-0) : 5$  là 2.

- Theo quan niệm thông thường, các điểm 1-2 được coi là kém, 3-4 là yếu, 5-6 là trung bình, 7-8 là khá, 9-10 là giỏi. Để phù hợp với quan niệm đó, ta có thể chia đoạn  $[0 ; 10]$  thành 4 nửa khoảng và 1 đoạn như sau:

$[0 ; 2,5), [2,5 ; 4,5), [4,5 ; 6,5), [6,5 ; 8,5), [8,5 ; 10]$ .

- Trong cả 5 lớp số liệu được phân ra, không lớp nào có tần số bằng 0, vậy ta khẳng định sự chia lớp theo 4 nửa khoảng và 1 đoạn được kí hiệu như trên, không phải điều chỉnh gì.

**Bước 3. Lập bảng tần số và tần suất của các lớp**

Lớp	$[0 ; 2,5)$	$[2,5 ; 4,5)$	$[4,5 ; 6,5)$	$[6,5 ; 8,5)$	$[8,5 ; 10]$	Tổng
Tần số	4	21	21	5	2	53
Tần suất	7,5%	39,6%	39,6%	9,4%	3,8%	100%

**2. Vẽ biểu đồ tần suất hình quạt**

Theo chuẩn kiến thức, kĩ năng của nội dung này trong môn Toán cấp THPT, chương trình chuẩn chỉ yêu cầu học sinh đọc được chứ không yêu cầu vẽ biểu đồ tần suất hình quạt, còn chương trình nâng cao cũng chỉ yêu cầu học sinh “vẽ được biểu đồ tần suất hình quạt trong trường hợp đơn giản” ([1], tr. 108 và tr. 177). Không nên e dè đối với yêu cầu về công việc này bởi vì, một mặt kĩ năng vẽ biểu đồ tần suất hình quạt rất cần trong đời sống để biểu thị trực quan những thông tin về cơ cấu thành phần, và mặt khác kĩ năng này chỉ đòi hỏi những kiến thức và kĩ năng cơ bản về đường tròn và góc ở tâm mà học sinh đã được học kĩ ở lớp 9. Và lại, vẽ biểu đồ tần suất hình quạt thì những trường hợp đơn giản và những trường hợp phức tạp cũng không phân biệt nhau bao nhiêu.

Vì vậy, khi đổi mới chương trình vào thời điểm tới gần đây nhất [2], trong chuẩn kĩ năng, yêu cầu vẽ được biểu đồ tần suất hình quạt nên được nêu lên không chỉ trong chương trình nâng cao mà còn cả trong chương trình chuẩn. Ngay ở thời điểm hiện tại, cách vẽ biểu đồ này cũng có thể được trình bày thông qua một bài đọc thêm hoặc một chủ đề nâng cao của môn Toán lớp 10.

Sau đây là hai cách vẽ biểu đồ tần suất hình quạt: cách vẽ nhờ máy vi tính và cách vẽ thủ công được trình bày thông qua một ví dụ.

*Ví dụ: Hãy vẽ biểu đồ hình quạt biểu diễn kết quả thi môn Toán học kì 2 của một lớp 10 THPT đã được nêu ở ví dụ 2 của mục 1.*

**Cách 1: Dùng máy vi tính cùng với phần mềm Excel**

Trong Excel, hãy nhập bảng phân bố tần suất ghép lớp (ví dụ 2 của mục 1) hoặc bảng chỉ gồm hai dòng tên các lớp và tần suất, chọn bảng này rồi lần lượt dùng các lệnh trong dòng lệnh dưới đây:

Insert\chart\Pie\chart sub-type\Serie in

Ở cửa sổ Serie in chọn Rows hay Column tùy theo dãy dữ liệu được nhập vào theo dòng hoặc cột. Tiếp đó chọn Next thì hình vẽ biểu đồ hình quạt hiện ra trên màn hình.

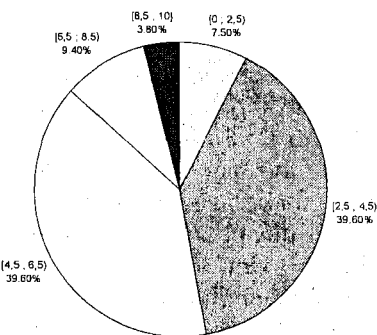
Khi đó, làm theo các chỉ dẫn trên màn hình để ghi tên biểu đồ, chú giải, tên của các lớp và những số phần trăm.

Cuối cùng chọn Finish để kết thúc rồi lưu trữ biểu đồ nhờ dòng lệnh

File\Save As

Bảng phân bố tần suất ghép lớp điểm thi môn Toán học kì 2 của một lớp 10 THPT năm học 2007-08

Lớp	[0 ; 2,5)	[2,5 ; 4,5)	[4,5 ; 6,5)	[6,5 ; 8,5)	[8,5 ; 10]	Tổng
Tần số	4	21	21	5	2	53
Tần suất	7,5%	39,6%	39,6%	9,4%	3,8%	100%
Góc h.quạt	27°	142,6°	142,6°	33,9°	13,7°	360°



**Cách 2: Dùng thước đo góc vẽ bằng tay**

a) Lập bảng giá trị các góc của các hình quạt biểu diễn phân bố tần suất ghép lớp

Trước hết ta tính số đo (độ) của các góc của các hình quạt biểu diễn các tần suất 7,5%, 39,6%, 39,6%, 9,4% và 3,8%.

Toàn bộ số liệu thống kê ứng với toàn bộ hình tròn tức là góc 360°. Vậy 7,5% số liệu ứng với góc  $360^\circ \times 0,075 = 27^\circ$ ,

39,6% số liệu ứng với góc  $360^\circ \times 0,396 = 142,56^\circ \sim 142,6^\circ$ ,

9,4% số liệu ứng với góc  $360^\circ \times 0,094 = 33,84^\circ \sim 33,9^\circ$ ,

3,8% số liệu ứng với góc  $360^\circ \times 0,038 = 13,68^\circ \sim 13,7^\circ$ .

Như vậy ta có thể bổ sung một dòng ghi các góc của các hình quạt biểu diễn các tần suất vào cuối bảng phân bố ghép lớp ở ví dụ 2 của mục 1 để được bảng sau:

b) Vẽ biểu đồ hình quạt biểu diễn phân bố tần suất ghép lớp

Vẽ một đường tròn rồi dùng thước đo góc vẽ các hình quạt liên tiếp kề nhau có số đo (độ) của góc ở tâm tương ứng lần lượt là 27°, 142,6°, 142,6°, 33,9° và 13,7°, ta được biểu đồ hình quạt biểu diễn phân bố tần suất ghép lớp đã biết.

**TÀI LIỆU THAM KHẢO**

1. Chương trình Giáo dục phổ thông cấp trung học phổ thông (2006), NXB Giáo dục.
2. Dự thảo chiến lược giáo dục 2009–2020, <http://vietnamnet.vn/giaoduc/2008/12/819333/> ngày 18/12/2008.

**SUMMARY**

Probability and statistics knowledge which have been introduced into the mathematics curriculum of the upper secondary school have potential to be applied in the real life. But this potential has been sometime wasted because the application of this knowledge needs rather complicated processes. Therefore this article is intended to help upper secondary teachers to overcome this difficulty so that the above mentioned potential can be exploited in some special cases: teaching to set up class frequency - relative class frequency distribution tables and teaching to draw relative frequency pie charts.