

ỨNG DỤNG MỘT SỐ PHƯƠNG PHÁP TOÁN TRONG NGHIÊN CỨU Y- SINH HỌC Ở NƯỚC TA

Vương Văn Hùng

Hiện nay, trong nghiên cứu y-sinh học có sử dụng các phương pháp toán thống kê một cách khá phổ biến để trình bày các kết quả nghiên cứu và đây cũng là một yêu cầu quan trọng khi đánh giá chất lượng của công trình nghiên cứu. Trong các phương pháp toán thì phổ biến nhất là sử dụng các số bình quân (số bình quân đặc trưng, số bình quân kết cấu)

Tuy nhiên, việc sử dụng các dạng số bình quân còn nhiều điều cần thảo luận.

Để góp phần giúp các bộ phận nghiên cứu sử dụng các phương pháp toán thống kê vào các kết quả nghiên cứu một cách thuận tiện và có tác dụng tốt xin nêu một số cách tính số bình quân thường dùng.

1. Số bình quân chung:

Công thức chung để tính số bình quân đặc trưng: $\bar{X} = \sqrt[k]{\frac{\sum x_i^k}{n}}$

\bar{X} Số bình quân bậc k, x_i giá trị biến thiên

k : Qui định dạng số bình quân

k = 1 số bình quân số học (áp dụng biến động số lượng trong giới hạn, các phân phối bậc 1). $\bar{X} = \frac{1}{n} \sum X_i$

2. Số bình quân kết cấu

2.1. Số bình quân bách phân vị (\bar{X}_p)

Được tính theo công thức $\bar{X}_p = \sqrt[p]{P_{X_i}} + d \cdot \frac{f_p - \sum f_{(i-1)}}{f_i}$

Trong đó: x_i trị số quan sát ở đơn vị i
 x_p số dưới của lớp có chứa (\bar{X}_p)

f_p : tần số p (%)

f_i : tần số lớp có chứa X_p

f_{i-1} : tần số lớp ngay trước lớp chứa X_p

d : khoảng cách các lớp.

2.2. Số trung vị (Mediale). Số đứng giữa dãy phân phối hoặc số trội (Mode) giá trị tại số có tần suất cao nhất (số phổ biến) và các số bình quân kết cấu khác.

Với các f_p được tính theo phần trăm, tính các giá trị đạt tới mức phần trăm ta đánh giá được kết cấu phân phối (đối xứng, chuẩn hay không) độ lệch, các mức giới hạn cao, thấp... trong mẫu nghiên cứu.

Bằng các bình quân tại các mức phần trăm ta có thể thay thế các số TB đặc trưng, đặc biệt nhóm phân phối số lượng còn nhỏ, chưa chuẩn, các mẫu phức tạp, mối quan hệ giữa các biến số dãy phân phối không phải bậc 1, ngay cả số TB ít có ý nghĩa đánh giá mẫu. Trong mẫu chuẩn $X_p = K_p \cdot \delta + X$ (δ : độ lệch chuẩn, K_p : hệ số từ hàm tích phân của phân phối chuẩn – có bảng tra cứu).

3. Ví dụ minh họa: Tính \bar{X}_p ở mức 5%, 50%, 95% của bảng số liệu sau

x_i	f_i	Tần số tích lũy (f_i)
45 – 54	10	10
55 – 64	30	40
65 – 74	45	85
75 – 84	80	165
85 – 94	30	195
95 – 104	5	200

Với $f_p 5\%$, $X_{5\%}$: thuộc vị trí $\frac{5 \times 200}{100} = 10$ cuối lớp đầu tiên (Giá trị giữa lớp 45 + d/2 là 50 thêm d/2) $X_{5\%} = 55$ (khoảng cách lớp d=10).

$$f_p 50\% = 50\% \times 200 : 100 \text{ thuộc lớp } \sum f_i = 165 \sum f_{(i-1)} = 85$$

$X_p = 75$ giá trị đầy lớp chứa f_d , $d = 10$.

Áp dụng công thức

$$X_{50\%} = 75 + 10 \times \frac{100 - 85}{80} = 76,875$$

$$\text{Tương tự } f_{95\%} = 200 \times 95 : 190$$

X_{95} thuộc lớp f_i bằng 195, $\sum f_i - 2 = 165$
 $f_i = 30$, $x_p = 85$

$$\text{Vậy } X_{95} = 85 + 10 \times \frac{190 - 85}{80} = 93,333$$

Nếu tính số TB đặc trưng đơn thuần ta có:

$\bar{X} = 75,25$ và $\delta = 11,3$ nếu số liệu đủ lớn ta suy ra mức giới hạn 5%, 50%, 95% như sau (tra bảng $K_p = 1,645$ mức 95%, 000 mức 50%, -1,645 mức 5%):

X_p ở mức 5%:

$$75,25 + 11,33 \times (-1,645) = 56,662$$

X_p ở mức 95%:

$$75,25 + 11,33 \times 1,645 = 93,838$$

X ở mức 50% tương đương:

$$\bar{X} (K_p : 000)$$

So sánh 2 cách tính bách phân vị ta có thể dùng được trong điều kiện số lượng nghiên cứu đủ lớn (thống kê), nhưng số lượng (n) nhỏ các giá trị sẽ sai lệch với giá trị thực tế khá lớn.

4. Một số điểm cần lưu ý khi sử dụng số bình quân trong nghiên cứu y – sinh học

- Số bình quân là một thống kê mô tả quan trọng, nhưng nó chỉ có tác dụng khi được tính toán trên tổng thể mẫu đủ đại diện. Trong trường hợp ngược lại, thông số này sẽ không có ý nghĩa thực tiễn;

- Tùy theo tính chất của nguồn số liệu thu thập được sẽ chọn công thức tính số bình quân thích hợp;

- Trước lúc tính các số bình quân phải sử dụng thêm các phương pháp khác để làm sạch số liệu chẳng hạn như quy tắc 3 xích ma (3σ) để loại bỏ các hiện tượng ngoại lai;

- Sau khi tính ra các số bình quân, phải tính toán phạm vi sai số thông qua chỉ tiêu độ lệch tiêu chuẩn.

Trên đây là một số ý kiến về sử dụng các số bình quân thông dụng thường gặp trong nghiên cứu y – sinh học và các lưu ý cần quan tâm trong khi sử dụng các số bình quân ■