ĐÁNH GIÁ CHẤT LƯỢNG CÁC BẢN SỐ LIỆU THÁM KHÔNG CỦA HÀ NỘI 1962 - 1977 BẰNG PHƯƠNG PHÁP ĐIỀU TRA CHỌN MẪU

Hoàng Thế Xương
(Đài Cao không - TU)

I- Đặt vấn đề :

Trong gần 20 năm nay số liệu thám không Hà nội đã tích lũy được khá nhiều, nhưng việc kiểm tra đánh giá chất lượng các bản số liệu này gặp nhiều khó khăn nên không tiến hành được kịp thời và đầy đủ. Hiện nay các số liệu đó cần phải được đánh giá chất lượng để phục vụ được rộng rãi và có hiệu quả. Hiện nay ta chưa có điều kiện thuận lợi để tiến hành kiểm tra số liệu bằng máy tính điện tử, đồng thời cũng không thể dùng phương pháp thủ công để tổ chức kiểm tra toàn bộ các bản số liệu trong một thời gian ngắn. Do đó dùng phương pháp điều tra chọn mẫu là một trong những vấn đề cơ bản của toán học thống kê được ứng dụng rộng rãi trong công nghiệp, nông nghiệp v.v. để kiểm tra đánh giá chất lượng sản phẩm, điều tra sản lượng, năng suất, v.v.

Phương pháp này có ưu điểm là tiến hành nhanh chóng, tiết kiệm thời gian và tiền của, tạo khả năng giúp ta đi sâu nghiên cứu tỷ mỷ từng phần tử (hay còn gọi là đơn vị) điều tra, từ đó giúp ta thu thập tài liệu một cách rộng rãi và chính xác.

II- Tóm tắt nội dung của phương pháp điều tra chọn mẫu :

A - Từ tập hợp tổng quát gồm N phần tử (tức là kích thước của tập hợp tổng quát gồm N đơn vị) ta lấy ra ngẫu nhiên n phần tử (tức là mẫu có kích thước n đơn vị) sau đó tiến hành quan sát n phần tử mấy rồi căn cứ vào kết quả thu được ta suy kết quả cho toàn tập hợp tổng quát.

Phương pháp này có những ưu điểm như đã nêu trên và tất nhiên nó có sai lệch so với phương pháp điều tra toàn diện. Sai lệch này phụ thuộc vào phương pháp chọn mẫu và kích thước n của mẫu.

Phương pháp chọn mẫu chỉ có tính đại diện cao khi :

1. Qúa trình sản xuất ra các phần tử điều tra ổn định, điều đó đảm bảo tính đồng nhất của các phần tử.

2. Các phần tử lấy ra từ tập hợp tổng quát để lập mẫu là những đại lượng ngẫu nhiên.

3. Số lượng hay kích thước của mẫu càng nhiều thì kết qủa thu được của mẫu sẽ càng gần sát với kết qủa của tập hợp tổng qúat,

Những yêu cầu đó cho phép sử dụng các kết qủa của lý thuyết xác suất vào điều tra chọn mẫu.

B- Giả sử tập hợp tổng qúat kích thước N có M đơn vị mang đấu hiệu A nào đó mà ta cần nghiên cứu. Số lượng M đơn vị đó (hoặc tỷ lệ $\frac{M}{N}$) ta chưa biết và cần phải xác định. Thí dụ trong 1 lô N sản phẩm có M phế phẩm (đấu hiệu A ở đây là phế phẩm). Số lượng M phế phẩm đó (hoặc tỷ lệ $p = \frac{M}{N}$) cần phải xác định. Vì N qúa lớn, không thể quan sát hết được nên phải điều tra bằng mẫu. Muốn vậy ta sẽ lấy một mẫu ngẫu nhiên gồm n sản phẩm chẳng hạn. Giả sử trong mẫu lấy ra có m phế phẩm, khi đó ta sẽ coi " xấp xỉ " tỷ lệ phế phẩm $\frac{m}{n}$ của mẫu cũng là tỷ lệ phế phẩm $p = \frac{M}{N}$ của tập hợp tổng qúat, tất nhiên có một sai lệch.Tùytheo mức độ chính xác đòi hỏi và tùy theo hoàn cảnh ta có thể qui định sai lệch này không vượt qúa một số ε cho trước, tức là ta sẽ khống chế :

$$\left|\frac{m}{n} - p\right| \leqslant \varepsilon \qquad (1)$$

Số ε gọi là sai số cho phép hoặc độ sai lệch của ước lượng.

Từ (1) biết tỷ lệ $\frac{m}{n}$ của mẫu và căn cứ vào sai số cho phép ta có thể nói tỷ lệ p của tập hợp tổng qúat sẽ là một giá trị nằm trong khoảng tin cậy $\frac{m}{n} - \varepsilon$ và $\frac{m}{n} + \varepsilon$. Nhưng rõ ràng kết luận này chưa hoàn toàn đúng vì ở đây ta ước lượng, nên kết luận chỉ có tính chất tương đối. Vì vậy ta cần đánh giá xem kết luận của ta đáng tin cậy tới mức độ nào. Nghĩa là tìm xác suất của biến cố $\left|\frac{m}{n} - p\right| \leqslant \varepsilon$. Người ta thường ký hiệu xác suất $P\left(\left|\frac{m}{n} - p\right| \leqslant \varepsilon\right) = \gamma$

γ được gọi là xác suất tin cậy của ước lượng.

C - Cơ sở lý luận của ước lượng này là qui luật phân phối nhị thức và định luật số lớn của Béc-nu-li.

$$\lim_{n \to \infty} P\left(\left|\frac{m}{n} - p\right| \leqslant \varepsilon\right) = 1 \qquad (2)$$

Từ (2) Béc-nu-li cho ta biết nếu số phép thử n rất lớn thì việc tần suất $\frac{m}{n}$ của một biến cố sai khác với xác suất p của nó một lượng nhỏ hơn một số ε tùy ý cho trước là một sự kiện gần như chắc chắc.

Như thế việc lấy tỷ lệ $\frac{m}{n}$ của mẫu để ước lượng tỷ lệ p của tập hợp tổng qúat là hợp lý. Nhưng điều tra chọn mẫu thì n không thể lớn vô cùng mà là một số hữu hạn, Khi đó ta sử dụng công thức xấp xỉ :

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) \qquad (3)$$

trong đó $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$ là tích phân xác suất và là hàm số $t = \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}$

$$t = \frac{\varepsilon\sqrt{n}}{\sqrt{pq}} \qquad (4)$$

Giá trị của $\Phi(t)$ được tính sẵn ứng với từng giá trị cụ thể của t theo (4) và lập thành bảng tra cho ở cuối các sách thống kê.

Từ (3) ta thấy xác suất tin cậy γ của ước lượng chính là $2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right)$

Từ (4) ta dễ dàng tìm được kích thước của mẫu ngẫu nhiên cần lấy ra từ tập hợp tổng quát với các giá trị cho trước của hai trong ba điều kiện :

- Xác suất p của biến cố
- Độ sai lệch ε của ước lượng.
- Độ tin cậy γ của ước lượng

Ta thấy n tỷ lệ thuận với pq. Vì tổng 2 số p và q luôn luôn bằng 1 nên tích của pq lớn nhất khi p = q = 0,5 là 0,25.

III - Ứng dụng phương pháp điều tra chọn mẫu ngẫu nhiên để đánh giá chất lượng các bản số liệu thám không.

A - Ở đây đã ứng dụng phương pháp điều tra chọn mẫu ngẫu nhiên nêu ở II để đánh giá chất lượng các bản số liệu thám không.

1. Vì chỉ đánh giá chất lượng các bản số liệu thám không, chứ không đánh giá các yếu tố khí tượng cao không, nên các bản số liệu ấy là 1 tập hợp các đại lượng ngẫu nhiên.

2. Các bản số liệu thám không Hà nội trong thời gian qua được thu thập, quai toán và thành lập trong một điều kiện ổn định và đồng nhất : cùng hệ máy Ma - lu-Khít - A_{22} trên cùng một địa điểm và theo một qui phạm kỹ thuật thống nhất.

3. Căn cứ vào yêu cầu phục vụ trước mắt và lâu dài, vào hoàn cảnh cụ thể hiện tại chúng tôi đặt yêu cầu độ sai lệch ε của ước lượng xác suất của biến cố

là 0,05 và độ tin cậy γ của ước lượng là 0,9999. Từ đó tính được số n lớn nhất cần chọn ra để lập một mẫu ngẫu nhiên là 1540 đơn vị.

4. Đánh số tập hợp các bản số liệu trong 15 năm theo thứ tự từ 00.000cho đến hết. Mở ngẫu nhiên các bảng số ngẫu nhiên Ka-đơ-rốp và theo phương phápchọn mẫu ngẫu nhiên lặp lại để chọn 1540 đơn vị có số hiệu trong phạm vi các số hiệu của tập hợp tổng quát. Đối chiếu các số hiệu chọn được bằng cách ngẫu nhiên ấy với bảng số thứ tự của tập hợp tổng quát, ta tìm được các bản số liệu cụ thể phải chọn.

Kết qủa các bản số liệu chọn được phân phối khá chuẩn trong các năm.

5. Lập một nhóm kiểm soát viên nắm vững phương pháp kiểm soát và cáctiêu chuẩn đánh giá chất lượng các bản số liệu theo qui phạm quan trắc VTTK. Kết qủa kiểm tra và nhận xét về từng bản số liệu được ghi vào nhật ký công tác của các kiểm soát viên.

B. Khi đặt vấn đề đánh giá chất lượng các bản số liệu, cần quan tâm đến những sai sót có ý nghĩa căn bản mà từ đó làm cho số liệu kém tin cậy hoặc dẫn đến nhiều sai sót khác. Ở đây chia ra làm 10 nhóm, mỗi nhóm đặc trưng cho mức độ và ý nghĩa khác nhau. Ví dụ nhóm I có ý nghĩa rất quan trọng, nó cho biết tỷ lệ các bản số liệu có được do các máy thám không đủ tiêu chuẩn, nhóm III có ý nghĩa cho các nhà nghiên cứu khi muốn sử dụng số liệu các tầng đặc tính để dựng lại một mặt cắt thẳng đứng của một lần thả máy nào đó, nhóm VIII cho biết ngoài tỷ lệ sai sót đã biết qua nhóm I còn có thể tìm thấy trong nhóm VIII, v.v

Dựa vào nhật ký công tác của các kiểm soát viên để tổng kết và đánh giá kết qủa kiểm tra tệp mẫu và tỷ lệ sai sót của các nhóm được ghi ở bảng 1.

Từ bảng 1 ta có những nhận xét sau đây :

1. Các năm đầu thả máy A_{22} đã thả nhiều máy không đủ tiêu chuẩn, có thể nói năm 1962 có đến gần 1/4, năm 1963 tới 1/10 và trong tập hợp mẫu có 2,65% số bản số liệu do các máy không đủ tiêu chuẩn cung cấp. Cũng trong những năm đầu ấy đã nhầm lẫn giữa 2 lần hiệu chuẩn nhiệt độ cao và nhiệt độ thấp : các năm 1962, 1963, 1964 đều có tới 1/3 số máy thả bị hiệu chuẩn nhầm lẫn.

2. Chất lượng qui tóan và thành lập các bản số liệu chứa nhiều sai sót căn bản là vào vài năm đầu thả máy A_{22} và vào các thời kỳ quan trắc viên mớivào nghề.

3. Chất lượng qui toán và thành lập các bản số liệu hầu như không chứa những sai sót căn bản vào những năm 1964, 1965, 1966, 1968 và 1972.

4. Trong 10 nhóm sai sót đáng kể nhất là nhóm I và nhóm VIII. Sai về độ

cao do mọi nguyên nhân dẫn đến, cũng chỉ chiếm 1,54% (cộng nhóm II và nhóm VII). Từ nhận xét này đã cho phép chúng tôi triểnkhai tiếp biện pháp kiểm tra để làm sạch toàn bộ sai sót nhóm I và nhóm VIII của tập hợp tổng quát. Biện pháp kiểm tra này đã đạt kết qủa loại ra hết các bản số liệu do máy thả không đủ tiêuchuẩn

5. Các nhóm sai sót khác không đáng kể (chưa đến 1%).
6. Chất lượng các bản số liệu trong tập mẫu :

Về Temp sau khi tính đến các sai sót nhóm I,II,III,VII,IX và Xcchất lượng các bản số liệu tốt đạt 95,4%.

Về gió trên cao sau khi tính đến các sai sót nhóm IV, V và VI chất lượng các bản số liệu gió đạt 98,51%.

Vì số lượng các bản số liệu Temp và gió đều bằng 1540 bản, tức là trọng lượng tham gia vào tính,chất lượng trung bình của các bản số liệu thám không đều bằng nhau, nên chất lượng chung của chúng à 96,96%.

Theo điểm 3 của phần A (III) trong tập mẫu đã thu được xác suất của biến cố các bản số liệu tốt là 96,96%. Theo phần B (II) ta có xác suất các bản số liệu tốt trong tập hợp tổng quát (tập hợp với các bản số liệu trong 15 năm) là một giá trị nằm trong khoảng từ 91,96 - 100% với độ tin cậy là 99,99%.

IV - Kết luận :

Bằng phương pháp điều tra chọn mẫu chỉ trong một năm kết qủa thu được như sau :

1. Đánh giá được mức độ các sai sót chứa trong tập hợp tổng quát các bản số liệu thám không Hà nội 15 năm (xem bảng 1).

2. Đánh giá được chất lượng của tập hợp tổng quát các bản số liệu ấy,Chất lượng này (96,96% ± 5%) đáp ứng được yêu cầu sử dụng với độ tin cậy cócủa ước lượng là 99,99%.

3. Loại các bản số liệu đo máy thám không đủ tiêu chuẩn đo ra khỏitập hợp các bản số liệu 15 nămqua, làm cho tập hợp ấy chỉ chứa những bản số liệu đo được bởi các máy thám không tốt.

Tuy nhiên phương pháp điều tra chọn mẫu mới chỉ chophép chúng tôi đánh giá chất lượng chung của tập hợp tổng quát các bản số liệu trong 15 năm và ở chừng mực nhất định cho phép đánh giá chất lượng của từng năm quan trắc.Việcđánh giá chất lượng theo mùa (mùa hè và mùa đông)của tập hợp tổng quát này , chưa giải quyết được.

(xem tiếp trang 36)

- Thời gian truyền lũ từ Hưng thi về Bến đế ngắn = 10h, (= 18-24h).

- Thời gian nước tập lưu nhanh, lượng mưa ngày lớn nhất trong ngày và đỉnh lũ cũng hình thành trong ngày 22.

- Thời gian nước lên nhanh, từ chân đến đỉnh lũ chỉ trong vòng 20h.

- Cường suất lũ lớn (Hưng thi 173cm/h, Bến đế 44cm/h). (xem bảng 4)

- Biên độ lũ lớn tại Hưng thi biên độ 1120cm, tại Bến đế biên độ 325cm).

Tóm lại năm 1978 lũ sông Hoàng long thuộc loại lũ lịch sử và nó có những đặc điểm riêng của nó. Trong bài này không trình bày các phương pháp hoàn nguyên lũ. Phần này trong phần tổng kết hoàn nguyên lũ năm 1978 của Cục dự báo khí tượng thủy văn đã báo cáo ở Hội đồng khoa học kỹ thuật Cục.

ĐÁNH GIÁ CHẤT LƯỢNG .....

(tiếp theo trang 35)

Tài liệu tham khảo

Tài liệu tham khảo

1. Lý thuyết xác suất và thống kê toán học của A.U.Ka-ra-xep, nhà xuất bản thống kê Mạc tư khoa, 1970.

2. Toán học cao cấp tập V của A-na-pê-trốp -va, Den-nep-va Ven-ep t-va, Đi-mô-va, Nan-tre-va, Ni-cô-lai-stôi-an-nốp. Nhà xuất bản kỹ thuật Xôphia - 1974.

3. Các bảng số ngẫu nhiên - Ka - đơ - rốp (Thư viện Tổng cục khí tượng thủy văn số TBS - 13).