Thử kiểm nghiệm sự phù hợp

CỦA TRUNG BÌNH LƯỢNG MƯA NĂM GIỮA CÁC THỜI KỲ BỊ GIÁN ĐOẠN BẰNG PHƯƠNG PHÁP KIỂM NGHIỆM THAM SỐ

Nguyễn Viết Lành - Đài KTTV Quảng ninh

I - Đặt vấn đề

Chuỗi số liệu khí tượng mà ta thu nhận được tại các trạm khí tượng trong quá trình quan trắc thường bị gián đoạn, hoặc về thời gian, hoặc về không gian. Sự gián đoạn đó ảnh hưởng rất lớn đến việc sử dụng số liệu khi nghiên cứu. Khi khai thác mà gặp phải chuỗi số liệu gián đoạn, ta thường phân vân : nếu không bỏ qua sự gián đoạn ấy và chỉ lấy một thời kỳ liên tục nào đó thôi thì chuỗi số liệu đó sẽ ngắn, không đảm bảo độ tin cậy của chuẩn khí hậu ; còn nếu bỏ qua sự gián đoạn ấy và coi nó như một chuỗi số liệu liên tục thì liệu có mắc sai lầm gì không ?

Chính vì vậy, trong bài này, chúng tôi sẽ thử sử dụng phương pháp kiểm nghiệm tham số để kiểm nghiệm sự phù hợp của trung bình lượng mưa năm giữa các thời kỳ bị gián đoạn của trạm khí tượng Hòng gai và giữa trạm khí tượng Hòng gai với trạm khí tượng Bãi cháy (chuyển từ Hòng gai sang).

II - Phương pháp và số liệu

1. Phương pháp : Kiểm nghiệm tham số, mà cụ thể ở đây là kiểm nghiệm sự bằng nhau của kỳ vọng mẫu với kỳ vọng cho trước hoặc một số cho trước.

Giả sử ta có một đại lượng ngẫu nhiên ξ với n trị số quan sát x_1, x_2, ..., x_n có phương sai đã biết. Ta sẽ kiểm nghiệm sự bằng nhau giữa kỳ vọng μ của tập mẫu này với kỳ vọng μ_0 cho trước nào đó. Giả thiết H_0 được đặt ra là : $\mu = \mu_0$. Nhưng trong thực tế phải thay μ bằng trung bình số học $\bar{x}$ và μ_0 bằng trung bình số học $\bar{x}_0$ khác. Như vậy, giả thiết H_0 trở thành : $\bar{x} = \bar{x}_0$.

Thực chất của việc kiểm nghiệm là ta xét xem trị số $|\bar{x} - \bar{x}_0|$ có lớn đến mức đáng kể hay không ? Nếu nó lớn quá thì giả thiết H_0 bị bác bỏ và nếu nó không đáng kể thì giả thiết H_0 được chấp nhận. Mức đáng kể đó thể hiện qua giới hạn tin cậy ban đầu d. Trị số này được xác định cụ thể tùy theo yêu cầu kiểm nghiệm trong công tác nghiên cứu. Như vậy chỉ tiêu để phán đoán là :

$$\begin{cases} |\bar{x} - \bar{x}_0| \geq d & \text{- bác bỏ giả thiết } H_0. \\ |\bar{x} - \bar{x}_0| < d & \text{- chấp nhận giả thiết } H_0. \end{cases} \qquad (1)$$

Để đảm bảo xác suất xảy ra sai lầm loại 1 (bác bỏ H_0 trong khi H_0 đúng) tương đối nhỏ, người ta đặt điều kiện :

$$p\left(|\bar{x} - \bar{x}_0| \geq d\right) = \alpha \qquad (2)$$

$$\text{Hay } p\left(\frac{|\bar{x} - \bar{x}_0|}{\sigma/\sqrt{n}} \geq \frac{d}{\sigma/\sqrt{n}}\right) \equiv p\left(|u| \geq U_\alpha\right) = \alpha \qquad (3)$$

Với $\sigma = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2}$, $\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$

Do vậy, chỉ tiêu (1) trở thành

$$\begin{cases} |u| \geqslant U_\alpha & \text{- bác bỏ giả thiết } H_0 \\ |u| < U_\alpha & \text{- chấp nhận giả thiết } H_0. \end{cases} \qquad (4)$$

Từ tập mẫu đã có, ta dễ dàng xác định được giá trị $|U|$

Cụ thể : $|u| = \frac{|\bar{x} - \bar{x}_0|}{\sigma} . \sqrt{n}$. Nhưng muốn tìm được U_α để so sánh, ta phải xác định được luật phân phối của U.

Lý thuyết xác suất thống kê [2] đã chứng minh rằng : nếu n đủ lớn thì $\bar{x}$ có phân phối chuẩn và theo [1] thì tổng lượng mưa năm thường có phân phối chuẩn với kỳ vọng bằng a và phương sai bằng σ^2. Khi đó, giả sử giả thiết H_0 đúng, nghĩa là : $M\bar{x} = \bar{x} = \bar{x}_0$ và $\sqrt{D\bar{x}} = \sigma / \sqrt{n}$ thì u có phân phối chuẩn với kỳvọng bằng 0 và phương sai bằng 1, ký hiệu $u \in N(0, 1)$.

Như vậy : $p(|u| \geqslant U_\alpha) = P(u \leqslant -U_\alpha) + P(u \geqslant +U_\alpha) = \alpha$

hay $\alpha = \int_{-\infty}^{-U_\alpha} \frac{1}{2\pi} e^{-\frac{1}{2}t^2} dt + \int_{+U_\alpha}^{+\infty} \frac{1}{2\pi} e^{-\frac{1}{2}t^2} dt = \int_{U_\alpha} \frac{1}{2\pi} e^{-\frac{1}{2}t^2} dt.$

do đó : $\frac{\alpha}{2} = \int_{+U_\alpha}^{+\infty} \frac{1}{2\pi} e^{-\frac{1}{2}t^2} dt.$

Trong các bảng xác suất, người ta lại lập ra trị số của hàm $f = \int_{-\infty}^{x} \frac{1}{2\pi} . e^{-\frac{1}{2}t^2} dt$ hay cũng chính là $1 - \int_{x}^{+\infty} \frac{1}{2} e^{-\frac{1}{2}t^2} dt.$

do đó, ta có :

$$1 - \frac{\alpha}{2} = \int_{-\infty}^{x} \frac{1}{2\pi} e^{-\frac{1}{2}t^2} dt. \qquad (5)$$

Dựa vào biểu thức (5), với giá trị cụ thể của α, ta tra bảng được giá trị U_α tương ứng.

Tóm lại, toàn bộ quá trình kiểm nghiệm ở đây là : tính giá trị $|u|$, từ tìm $1 - \frac{\alpha}{2}$ rồi tra bảng hàm phân phối chuẩn để tìm U_α ; phán đoán giả thiết H_0 theo chỉ tiêu (4).

2. Số liệu : Số liệu về tổng lượng mưa năm tại trạm khí tượng Hồng gai có từ năm 1918 đến năm 1944 và từ năm 1950 đến năm 1974. Tại trạm khí tượng Bãi cháy có từ năm 1974 đến năm 1981.

III - Tính toán và kết quả

Bằng những số liệu trên, chúng tôi đã tiến hành kiểm nghiệm sự phù hợp của trung bình lượng mưa năm giữa 2 thời kỳ (1918-1944 và 1950-1974) của trạm khí tượng Hồng gai. Nếu phù hợp thì lẽ dĩ nhiên là chúng tôi coi cả 52 năm đó như một chuỗi số liệu liên tục và tiếp tục kiểm nghiệm sự phù hợp của 52 năm đó với 18 năm(1974-1981) của trạm Bãi cháy. Nếu không phù hợp thì chúng tôi sẽ kiểm nghiệm sự phù hợp của từng thời kỳ ở trạm Hồng gai với trạm Bãi cháy để tìm ra một thời kỳ phù hợp nào đó. nhằm kéo dài chuỗi số liệu giúp cho việc nghiên cứu về sau.

Như đã nói ở trên, tùy theo yêu cầu về độ chính xác của của chuỗi sốliệu mà ta sử dụng vào việc nghiên cứu, ta có thể định ra trị số của giới hạn tin cậy ban đầu d, có nghĩa là định ra xác suất mắc sai lầm α. Ở đây chúngtôi sử dụng hai trị số của α là 0,01 và 0,05.

Kết quả tính toán được dẫn ra trong bảng sau :

Các thời kỳ được kiểm nghiệm		$\Delta\bar{x}$	σ	$\lvert u \rvert$	Kết quả ứng với α khác nhau	
(1)	(2)				$\alpha=0,01 \rightarrow U_\alpha=2,58$	$\alpha=0,05 \rightarrow U_\alpha=1,96$
1918-1944	1950-1974	209,7	431,9	2,52	$\lvert u \rvert < U_\alpha$	$\lvert u \rvert > U_\alpha$
1918-1974	1974-1981	46,0	478,6	1,55	$\lvert u \rvert < U_\alpha$	$\lvert u \rvert > U_\alpha$
1918-1944	1974-1981	54,8	431,9	0,66	$\lvert u \rvert < U_\alpha$	$\lvert u \rvert > U_\alpha$
1950-1974	1974-1981	154,9	492,0	1,04	$\lvert u \rvert < U_\alpha$	$\lvert u \rvert > U_\alpha$

IV - Kết luận

Theo chỉ tiêu (4), với kết quả ở bảng trên, ta có thể rút những kết luận sau :

1. Với xác suất mắc sai lầm là 1% thì chuỗi số liệu 60 năm (1918-1981) tuy bị gián đoạn cả về thời gian và không gian nhưng được coi như một chuỗi số liệu liên tục.

2. Với xác suất mắc sai lầm là 5% thì một trong hai thời kỳ của Bằng giang (1918 - 1944 hoặc 1950 - 1974) ghép với chuỗi số liệu của Kỳ cùng được coi là liên tục, còn bản thân hai thời kỳ đó ghép với nhau lại không được xem là liên tục./.

MÙA MƯA LŨ VÀ CÔNG TÁC DỰ BÁO LŨ NĂM 1982 Ở LƯU VỰC SÔNG THÁI BÌNH

(Tiếp theo trang 21)

dự báo điểm uốn lũ bắt đầu lên hoặc lũ bắt đầu xuống, phương án dự báo đỉnh lũ, cách xử lý những trường hợp có lũ gia nhập lớn, trường hợp ảnh hưởng lũ sông Hồng v.v...

Cũng qua mùa lũ năm nay (1982) và mấy mùa lũ gần đây chúng ta thấy rõ rằng rằng, tình hình thông tin liên lạc điện báo mưa lũ ngày một kém cả về số lượng lẫn chất lượng, nhất là mấy trạm điện báo mưa lũ thuộc tỉnh Lạng sơn. Mặt khác trên toàn sông Thái bình việc đầu tư nghiên cứu biên tập phương án dự báo lũ sông Cầu, sông Thương còn quá ít và chưa hệ thống. Vì vậy cho nên trong những năm tới cần phải đầu tư thêm thời gian để nghiên cứu và có thể phối hợp cùng với bộphận nghiên cứu dự báo thủy văn để thử nghiệm, ứng dụng mô hình SSARR vào dự báo lũ sông Thái bình./.