

수문통계학의 기초(Ⅳ)

Introduction to Statistical Hydrology(Ⅳ)

허 준 행*

- I. 통계학의 기초(Basic Statistics)
- II. 빈도해석(비매개변수적 방법)(Nonparametric Frequency Analysis)
빈도해석(매개변수적 방법)(Parametric Frequency Analysis)
- III. 검정방법(Various Tests)
- IV. 자료의 경향 및 변동 측정방법(Detection of Changes and Trend in Data)
- V. 결측치보완 및 자료확충방법(Filling in Missing Data and Extension of Records)

지난 강좌까지는 여러 가지 통계값 등 통계학의 기초, 강우량 및 홍수량 같은 수문자료의 빈도해석 방법, 수문자료에 대한 통계분석을 실시할 때 많이 사용하는 무작위성 검정, 정규분포 검정, 적합도 검정방법에 대해서 설명하였으며, 본 강좌에서는 주어진 수문자료의 변동 및 경향 분석방법에 대하여 연자료 및 계절자료인 경우에 대하여 설명하기로 한다.

4. 자료의 경향 및 변동 측정방법

일반적으로 수문자료를 해석하거나 모형화할 때, 주어진 자료의 평균과 분산값이 정상적(stationary)이라고 가정하는 경우가 많다. 그러나 주어진 자료에 변동(change)이나 경향(trend)이 나타나는 경우에는 이러한 특성을 파악하고 해석을 하든지, 아니면 이러한 특성을 제거한 뒤 해석을 해야

한다. 본 강좌에서는 자료의 평균과 분산의 변동 및 경향을 파악할 수 있는 매개변수적 또는 비매개변수적인 방법들에 대하여 간단히 설명하기로 한다.

4.1 자료의 변동 및 경향

수문 및 수자원 자료에서 나타나는 자료의 변동 및 경향은 크게 자연적인 현상과 인간에 의한 변화에 의해 발생하게 된다. 자료의 변동 및 경향을 발생시키는 자연적인 현상의 예로는 산불, 화산 폭발, 사면붕괴 등이 있으며, 인간에 의한 원인으로서는 상류부 댐 건설, 토지이용 변화, 산림 개발, 도로 개발 등이 있다.

일반적으로 변동과 경향을 파악하는 절차는 예비적 자료해석(exploratory data analysis)과 확정적 자료해석(confirmatory data analysis)으로 분류될 수 있다(McLeod 등, 1983). 예비적 자료해석 방법은 대상 자료를 도시하거나 자료의 기본 통계값을 계산하여 분석하는 방법이며, 확정적 자료 해석 방법은 자료의 변동과 경향을 통계적인 방법을 이용하여 구하는 방법이다.

4.2 도시적인 해석

도시적인 방법은 예비적 자료 해석 기법으로 자료의 변화 및 경향을 파악하는 과정에서 제일 먼저 실시하는 단계이다. 예를 들어 시계열 자료인 경우 간단하게 도시할 수 있으며 이로부터 자료의 변화

* 연세대학교 토목공학과 조교수

및 경향을 쉽게 파악할 수 있다. 도시적인 해석 방법으로는 다음과 같은 방법들이 제안되고 있다 (Tukey, 1977 ; McLeod et al., 1983 ; Helsel과 Hirsch, 1992).

- 시계열 도시 (time series plot)
- 단일누가곡선 도시 (single mass plot)
- 이중누가곡선 도시 (double mass plot)
- 박스 도시 (box plot)
- S-S plot
- Q-Q plot

여기서 시계열 자료의 도시는 시간에 따라 자료를 나타내는 가장 간단한 방법이며, 단일누가곡선은 그림 상에서 직선으로 나타나는 경우 변화가 없

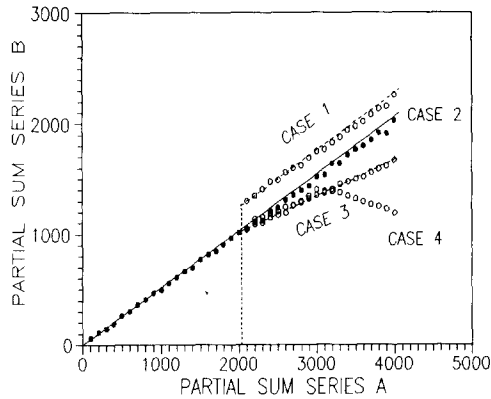


그림 4.1 이중누가곡선의 예

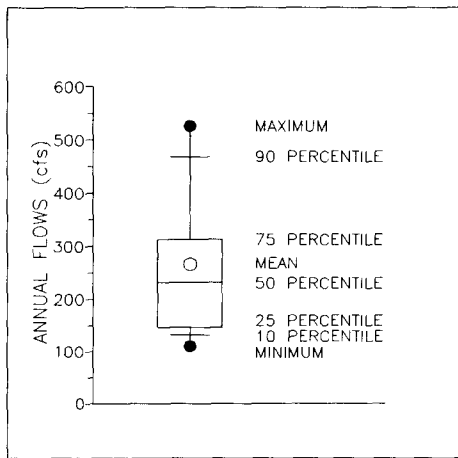


그림 4.2 박스형 도시 방법의 예

는 것을 쉽게 판단하는 방법이다. 이중누가곡선은 수문학에서 널리 사용하는 방법으로 그림 4.1에서 보는바와 같이 4가지 경우를 생각할 수 있다.

그림 4.1에서 보는바와 같이 자료에 변동이 없는 경우에는 이중누가곡선상에서 직선으로 표시되며 (CASE 2), 변동이 있을 때에는 CASE 3에 나타난 바와 같이 위 또는 아래로 경사지게 된다. 자료가 경향을 갖게 되는 경우에는 CASE 4에서와 같이 감소(또는 증가)하는 곡선으로 표시되고, 특이한 경우로 자료에 이상값(outlier)이 존재하는 경우에는 CASE 1과 같이 나타난다.

박스형 도시방법은 한눈에 자료의 중앙집중성향, 변화, 대칭성 및 극치값의 존재를 파악할 수 있는 방법으로 그림 4.2와 같이 표기된다(Helsel과 Hirsch, 1992).

4.3 평균 및 분산의 변화 측정 방법

본 절에서는 시계열자료의 평균과 분산의 변화를 파악하는 방법에 대해서 설명하기로 한다.

4.3.1 단일자료(single series)의 경우

단일자료의 평균과 분산의 변화 및 경향을 파악하는 방법으로는 Mann-Whitney(또는 Wilcoxon rank sum) test, sign test, Abbe test of homogeneity, simple T test, simple F test, modified T test, modified F test 등 외에도 많은 방법이 있으나 이중에서 몇 가지만 간단하게 소개하기로 한다.

- Mann-Whitney test

Wilcoxon(1945)이 처음으로 제안한 것을 Mann과 Whitney(1974)가 자료수가 다른 경우에도 적용할 수 있도록 변형시킨 방법으로 2개의 자료군의 분포가 동일하다는 것이 가정 사항이다. 자료수가 N 인 $Y_i (i=1, 2, \dots, N)$ 를 각각 N_1 및 N_2 의 자료수를 갖는 Y_1, Y_2, \dots, Y_{N_1} 과 $Y_{N_1+1}, Y_{N_1+2}, \dots, Y_N$ 으로 구분하고, 전체 자료를 크기순(작은 값부터 큰 값순으로)으로 정렬하여 변수 Z_i 로 정의한 뒤 첫 번째 자료의 평균과 두 번째 자료의 평균이 동일하다는 귀무가설 하에 다음의 검정량을

계산한다(Hollander와 Wolfe, 1973).

$$U_c = \frac{\sum_{i=1}^{N_1} (Y_i) - \frac{1}{2}N_1(N_1+N_2+1)}{\left[\frac{1}{12}N_1N_2(N_1+N_2+1) \right]^{1/2}} \quad (4.1)$$

여기서 $R(Y_i)$ 는 자료 Y_i 의 순서통계량의 서열(rank)로서, 두 개의 자료군의 평균값이 동일하다는 귀무가설은 다음과 같은 경우 성립한다.

$$|U_c| \leq u_{1-\alpha/2} \quad (4.2)$$

여기서 $u_{1-\alpha/2}$ 는 표준정규분포의 $1-\alpha/2$ 분위수(quantile)에 해당되는 값이며, α 는 유의수준이다.

• Sign test

Sign 방법은 대상 자료의 분포와 상관없이 사용할 수 있는 간단한 방법으로 자료수가 N 인 단일 자료를 자료수가 같은 2개의 자료군으로 나눈 뒤 다음과 같이 새로운 자료로 구분한다.

$$W_i = 1 \text{ if } Y_i < Y_{M+1} \quad (4.3a)$$

$$W_i = 0 \text{ if } Y_i \geq Y_{M+1} \quad (4.3b)$$

여기서 $M=N/2$ 이며 자료수 M 이 큰 경우($M > 20$) 검정량은 다음과 같이 주어지며 (Hollander와 Wolfe, 1973)

$$U_c = \left[\frac{2\sum_{i=1}^M W_i - M}{\sqrt{M}} \right] \quad (4.4)$$

2개의 자료군의 평균이 동일하다는 귀무가설은 다음의 식을 만족하면 성립한다.

$$|U_c| \leq u_{1-\alpha/2} \quad (4.5)$$

[예제 4.1] 나일강의 Aswan 지점의 1871~1944년까지의 연평균유출량은 그림 4.3과 같다($N=74$). 따라서 sign 방법에서의 $M=37$ 이다. 그림에

서 보는바와 같이 첫 번째 자료(1871~1907)의 평균값은 $102.686 \times 109m^3$, 두 번째 자료(1908~1944)의 평균값은 $83.197 \times 109m^3$, 74년 전 자료의 평균값은 $94.942 \times 109m^3$ 이다. 식 (4.3)과 (4.4)를 이용한 결과 $U_c=3.781$ 이며, 유의수준 $\alpha=0.05$ 에 대한 $u_{1-\alpha/2}=1.960$ 이므로 식 (4.5)에 의해 두 자료군이 동일한 평균값을 갖는다는 귀무가설은 기각된다.

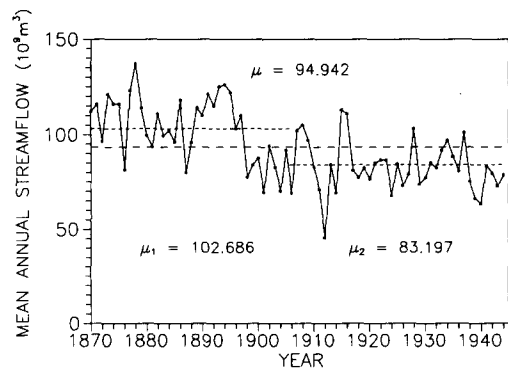


그림 4.3 나일강 Aswan 지점의 연평균유출량 (1871~1994).

4.3.2 다중자료(multiple series)의 경우

앞에서 제시한 단일자료의 평균 및 분산의 변화 및 경향을 분석하는 방법인 Abbe test, Mann-Whitney test, sign test, simple T test, simple F test, modified T test, modified F test는 다중 자료의 경우에도 적용할 수 있으며, 이외에도 이종누가곡선, Friedman test, Wilcoxon test, Kruskal-Wallis test, Barlett test, linear regression, analysis of variance test 등을 이용하여 다중자료의 변화 및 경향을 분석할 수 있다.

4.4 경향분석 방법

본 절에서는 주어진 연자료(annual series) 및 계절자료(seasonal series)의 평균과 분산의 경향을 파악하는 방법에 대하여 간단히 설명하기로 한다. 시계열 자료를 도시함으로써 자료의 경향을 쉽

게 파악할 수 있으나, 명확한 경향이 나타나지 않는 경우에는 통계적인 방법을 이용하여 경향을 분석할 수 있다. 참고적으로 연자료와 계절자료에 경향이 있는 경우에는 일반적으로 자료간에 강한 자기상관성을 갖는다.

4.4.1 연자료의 경우

연자료의 경향을 분석하는 방법으로는 T test, Hotelling-Pabst test, nonlinear trend test, Mann-Kendall test, Sen test 외에도 여러 가지 방법이 있으나 여기서는 언급한 방법 등에 대해 간단히 설명하기로 한다.

• T test

자료수 N인 연자료가 Y_1, Y_2, \dots, Y_N 으로 주어졌을 때 자료의 선형적 경향을 분석하는 방법으로 다음과 같은 선형방정식을 사용한다.

$$Y_t = a + bt \quad t=1, 2, \dots, N \quad (4.6)$$

여기서 주어진 자료가 선형적 경향을 갖고 있다면 계수 b는 0이 아닌 다른 값을 갖게 된다. 그러므로 기울기 b가 0인가 아닌가에 대한 귀무가설이 곧 주어진 자료에 대한 선형적 경향의 귀무가설이 된다.

[예제 4.2] 대관령의 연평균강수량(1978~1995)은 표 4.1과 같다. 대관령지점의 연평균강수량이 선형적 경향을 갖는지 알아보자. 강수량자료를 도시하면 그림 4.4와 같다.

표 4.1 대관령지점의 연평균강수량(1978~1995)
(단위 : mm)

연 도	연평균강수량	연 도	연평균강수량
1978	1605.2	1987	1626.7
1979	1193.0	1988	1968.1
1980	1626.4	1989	2194.0
1981	1529.4	1990	2316.2
1982	1252.6	1991	2461.5
1983	1270.3	1992	2187.9
1984	1699.3	1993	2353.6
1985	1744.7	1994	1444.1
1986	1476.1	1995	1337.8

대관령지점의 1978~1995년 연평균강수량에 대한 단순회귀분석(simple linear regression, Draper와 Smith, 1981) 결과 계수값은 $a=1371.362$, $b=38.610$ 이었다. 귀무가설 : $b=0$ 에 대한 검정 결과 기각되어 선형 경향이 있는 것으로 나타났다.

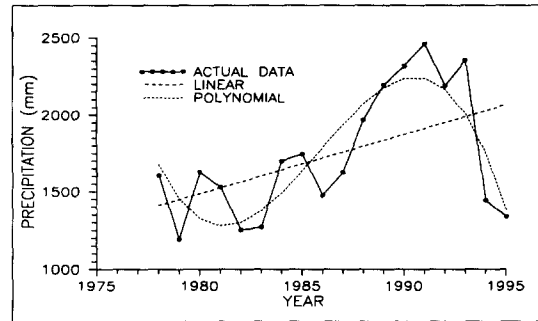


그림 4.4 대관령지점의 연평균강수량(1978~1995)

• Hotelling-Pabst Test

자료수 N인 연자료 Y_1, Y_2, \dots, Y_N 를 크기순으로(작은 값부터 큰 순서) 정리된 자료를 X_1, X_2, \dots, X_N 이라 하고 Y_i 의 순서값(rank)을 $R(Y_i)$ 라 하면 검정통계량은 다음과 같이 주어지고(Conover, 1971)

$$T = \sum_{i=1}^N [R(Y_i) - i]^2 \quad (4.7)$$

검정통계량 T의 상한계와 하한계는 아래와 같이 주어진다.

$$T_U = \frac{N(N^2-1)}{6} + u \frac{N(N^2-1)}{6\sqrt{N-1}} \quad (4.8a)$$

$$T_L = \frac{N(N^2-1)}{6} - u \frac{N(N^2-1)}{6\sqrt{N-1}} \quad (4.8b)$$

여기서 식 (4.7)에서 계산된 T값이 식 (4.8)밖에 존재하면 주어진 자료의 선형 경향은 유의수준 α 의 범위 내에서 받아들여진다고 할 수 있다.

[예제 4.3] 예제 4.2에서 주어진 대관령 자료에 대한 Hotelling-Pabst 계산 결과는 아래와 같다.

식 (4.7)로부터 $T=1404$ 이고, 식 (4.8)을 이용하여 $T_1=1429.6$, $T_2=508.4$ 이므로 유의수준 $\alpha=0.05$ 에서 T 값이 상한계와 하한계 안에 존재하므로 대관령 연평균자료는 선형적 경향을 갖지 않는 것으로 나타났다.

• 비선형 경향 분석방법

자료수 N 인 연자료 Y_1, Y_2, \dots, Y_N 가 주어졌을 때 Y_t 를 아래와 같은 H 차 다항식으로 표현할 수 있다고 하자.

$$Y_t = a + b_1 X_t + b_2 X_t^2 + \dots + b_H X_t^H \quad (4.9)$$

여기서 $X_t = t (t=1, 2, \dots, N)$ 로 주어지며 차수 H 와 계수를 구하는 방법으로는 all-, forward-, backward-, stepwise regression 등(Draper와 Smith, 1981)이 있으나 본 강좌에서는 생략하기로 한다. 자료의 분석결과 식 (4.9)에서 차수 H 가 2 이상인 경우에는 비선형 경향을 갖는다고 할 수 있다.

[예제 4.4] 대관령 강우자료에 대한 비선형 경향 분석 결과는 다음과 같다.

Backward 회귀분석 결과 결정된 차수는 $H=3$ 이었으며, 계수값은 다음과 같으므로 대관령 자료는 비선형 경향을 갖는다고 할 수 있다.

$$a=2007.939, b_1=-392.087,$$

$$b_3=61.821, b_3=-2.332$$

즉, $Y_t=2007.939-392.087t+61.821t^2-2.332t^3$ 이며 이에 대한 결과는 그림 4.4에서 polynomial로 표시된 곡선이다.

• Mann-Kendall test

연자료 $Y_t (t=1, 2, \dots, N)$ 에 대하여 $Y_{t'} (t'=1, 2, \dots, N-1)$ 과 $Y_t (t=t'+1, \dots, N)$ 의 크기를 비교하여 다음과 같이 정의한다.

$$Z_k = 1 \quad \text{if } Y_t > Y_{t'} \quad (4.10a)$$

$$Z_k = 0 \quad \text{if } Y_t = Y_{t'} \quad (4.10b)$$

$$Z_k = -1 \quad \text{if } Y_t < Y_{t'} \quad (4.10c)$$

여기서 $k=(t'-1)(2N-t')/2+(t-t')$ 라고 하면 Mann-Kendall 검정통계량은 다음과 같이 주어진다(Hirsch 등, 1982).

$$U_c = \frac{S+j}{[V(S)]^{1/2}} \quad (4.11)$$

여기서

$$S = \sum_{t'=1}^{N-1} \sum_{t=t'+1}^N Z_k \quad (4.12a)$$

$$V(S) = \frac{1}{18} \left[N(N-1)(2N+5) - \sum_{i=1}^G E_i(E_i-1)(2E_i+5) \right] \quad (4.12b)$$

$S < 1$ 이면 $j=1$, $S > 1$ 이면 $j=-1$, $S=0$ 면 $j=0$ 의 값을 가지며, G 는 같은 값을 갖는 자료군의 총수이며, E_i 는 i 번째 자료군에 속하는 같은 값을 갖는 자료의 수이다.

아래 식 (4.13)의 조건이 만족되면 유의수준 α 에서 주어진 자료가 상향(upward) 또는 하향 경향(downward trend)을 갖는다고 할 수 있다.

$$|U_c| > u_{1-\alpha/2} \quad (4.13)$$

주어진 자료에 대하여 $S=0$ 이면 $U_c=0$ 으로 경향을 갖지 않으며, S 가 양수인 경우에는 상향 경향을 가지며, S 가 음의 값을 갖는 경우에는 하향 경향을 갖는다.

[예제 4.5] 대관령 자료에 대한 Mann-Kendall 적용 결과는 아래와 같다.

유의수준 $\alpha=0.05$ 일 때 $u_{1-\alpha/2} = 1.96$ 이고, 식(4.11)로부터 $U_c=2.197$ 인 양의 값을 가지며 식(4.13)의 조건을 만족하므로 그림 4.4에서 보는바와 같이 상향 경향을 갖는다고 할 수 있다.

- Sen test

Sen (1968)에 의해 제안된 비매개변수적 방법으로, 주어진 연자료 $Y_i (i=1, 2, \dots, N)$ 에 대하여 다음과 같이 새로운 변수 Z_k 를 정의한다.

$$Z_k = \frac{Y_j - Y_i}{j - i} \quad k=1, 2, \dots, N(N-1)/2 \quad (4.14)$$

여기서 모든 i 와 j 에 대해 $j > i$ 이고, Z_k 를 다시 크기순으로 정렬한 값을 X_k 로 정의하면 Sen의 경사에 대한 검정통계량은 다음과 같이 주어진다 (Gilbert, 1987).

$$S = X_{(M+1)/2} \quad \text{if } M \text{ is odd} \quad (4.15a)$$

$$S = \frac{1}{2} [X_{M/2} + X_{(M+2)/2}] \quad \text{if } M \text{ is even} \quad (4.15b)$$

여기서 $M = N(N-1)/2$ 으로 식(4.15)에서 구한 검정통계량이 다음의 신뢰한계 안에 있게되면 유의수준 α 에서 경향을 갖는다고 할 수 있다.

$$[X_{M_L}, X_{M_U}] \quad (4.16)$$

여기서

$$M_L = \frac{M - u_{1-\alpha/2} [V(S)]^{1/2}}{2} \quad (4.17a)$$

$$M_U = 1 + \frac{M + u_{1-\alpha/2} [V(S)]^{1/2}}{2} \quad (4.17b)$$

이고 $V(S)$ 는 식 (4.12b)를 이용하여 구한다. 그러므로 하한계 X_{M_L} 은 X 값 중 가장 작은 값으로부터 M_L 번째 값이고, 상한계 X_{M_U} 는 M_U 번째 값이 된다.

[예제 4.6] 대관령자료에 대한 Sen 방법 결과는 아래와 같다.

자료수 $N=18$ (1978~1995)이므로 $M=153$ 이 되어 식(4.15a)를 사용하여 구한 $S=54.95$ 이다. $V(S)=697$ 이고 $u_{1-0.05/2}=1.96$ 이므로 식 (4.17)로부터 $M_L=50$, $M_U=103$ 이 된다. 따라서 하한계

$X_{M_L}=5.625$, 상한계 $X_{M_U}=89.1$ 의 값을 가지므로 대관령자료는 경향을 갖는다고 말할 수 있다.

4.4.2 계절자료의 경우

주어진 자료가 계절자료인 경우에 경향을 분석하는 방법으로는 자료평균의 선형 경향 분석을 위한 T test, 비선형 경향 분석법, seasonal Kendall test, 자료분산의 경향 분석방법 등 외에 여러 가지 방법이 있으나 본 강좌에서는 언급한 방법에 대해 간단히 설명하기로 한다.

- 평균의 선형적 경향을 위한 T test

계절자료 $Y_{\nu, \tau} (\tau=1, \dots, N \text{ 자료기간}; \nu=1, \dots, \omega \text{ 계절수})$ 가 주어졌을 때 연평균자료는 다음과 같이 정의된다.

$$Z_{\nu} = \frac{1}{\omega} \sum_{\tau=1}^{\omega} Y_{\nu, \tau} \quad (4.18)$$

식 (4.18)의 자료가 경향을 갖는다면 다음과 같은 선형회귀방정식으로 나타낼 수 있다.

$$Z_{\nu} = a + b\nu \quad (4.19)$$

여기서 기울기 b 의 값이 통계학적으로 0이 아닌 것을 확인해야 한다. 만약 Z_{ν} 가 경향을 갖는 경우 전체 계절자료를 이용한 회귀방정식은 다음과 같이 쓸 수 있다.

$$Y_{\nu, \tau} = a + b i \quad (4.20)$$

여기서 $i = (\nu-1)\omega + \tau$ 이다.

[예제 4.7] 대관령지점의 월평균 강수량 (1978~1995)은 그림 4.5와 같다. 월평균강수량자료에 대한 T test 결과는 아래와 같다.

귀무가설 : $b=0$ 에 대해서는 유의수준 $\alpha=0.05$ 에서 기각되었으며, 단순회귀분석 결과 $a=114.048$, $b=0.293$ 인 것으로 나타났다. 이에 대한 선형적 경향을 그림 4.5에 표시하였다.

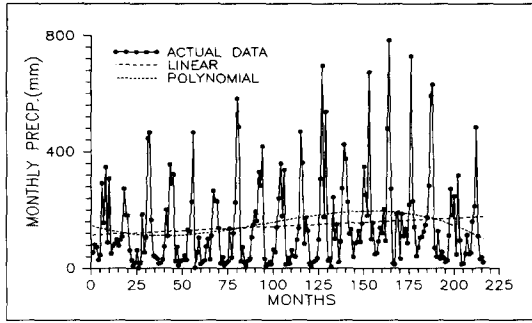


그림 4.5 대관령지점의 월평균강수량(1978~1995)

• 평균의 비선형 경향 분석

계절자료 $Y_{\nu, \tau}$ ($\tau=1, \dots, N$ 자료기간; $\nu=1, \dots, \omega$ 계절수)가 주어졌을 때 식 (4.21)과 (4.22)와 같이 새로운 변수를 정의하고

$$X_{\nu} = \nu \quad (4.21)$$

$$Z_{\nu} = \frac{1}{\omega} \sum_{\tau=1}^{\omega} Y_{\nu, \tau} \quad (4.22)$$

Z_{ν} 가 차수가 H인 비선형 경향을 갖는다면 다음과 같이 표기되며

$$Z_{\nu} = a + b_1 X_{\nu} + b_2 X_{\nu}^2 + \dots + b_H X_{\nu}^H \quad (4.23)$$

식 (4.23)에서 H가 1보다 큰 경우에 비선형 경향을 갖는다고 할 수 있다.

자료가 비선형 경향을 갖는 경우 전체 계절자료를 이용한 다항식으로 표현할 수 있다.

$$U_i = a + b_1 V_i + b_2 V_i^2 + \dots + b_H V_i^H \quad (4.24)$$

여기서 $U_i = Y_{\nu, \tau}$ 이고 $V_i = i$ 이며 $i = (\nu-1)\omega + \tau$ 이다.

[예제 4.8] 대관령지점의 월평균강수량자료에 대한 비선형 경향 분석 결과는 다음과 같다.

대관령자료를 이용하여 구한 식 (4.24)의 차수는 3이며, 다항식은 다음과 같이 주어지므로 월평균강수량은 비선형 경향을 가지며 그림 4.5에 나타내었다.

$$Y_{\nu, \tau} = 147.813 - 2.132i + 0.319i^2 - 0.001i^3$$

• Seasonal Kendall test

계절자료 $Y_{\nu, \tau}$ ($\tau=1, \dots, N$ 자료기간; $\nu=1, \dots, \omega$ 계절수)에 대하여 계절별로 다음과 같이 정의할 수 있다.

$$X_{\nu} = Y_{\nu, \tau} \quad \nu=1, \dots, N \quad (4.25)$$

또한 각 계절별로 Mann-Kendall 통계량 S_{τ} 와 $V(S_{\tau})$ 은 식 (4.12a)와 (4.12b)로부터 구할 수 있으며 이를 이용하여 다음과 같이 새로운 변수를 정의하면

$$SS = \sum_{\tau=1}^{\omega} S_{\tau} \quad (4.26a)$$

$$V(SS) = \sum_{\tau=1}^{\omega} V(S_{\tau}) \quad (4.26b)$$

식 (4.26)의 결과를 이용하여 검정통계량은 다음과 같이 정의된다.

$$U_c = \frac{SS + j}{[V(SS)]^{1/2}} \quad (4.27)$$

식 (4.28)의 조건이 만족되면 유의수준 α 에서 주어진 자료가 상향(upward) 또는 하향 경향(downward trend)을 갖는다고 할 수 있다.

$$|U_c| > u_{1-\alpha/2} \quad (4.28)$$

주어진 자료가 경향을 갖는 경우 seasonal Kendall 경사 (Sen test에서 설명)를 이용하여 경향을 파악할 수 있다. 식 (4.25)로부터 주어지는 자료에 대하여 다음과 같이 변수를 정의하고 ($k=1, \dots, M=N(N-1)/2$)

$$Z_{k, \tau} = \frac{X_t - X_i}{t-i} \text{ for } i, t=1, \dots, N \text{ and } t>1 \quad (4.29)$$

다시 식 (4.30)과 같이 정의한다.

$$X_j = Z_{k, \tau} \quad (4.30)$$

여기서 $j=1, \dots, \omega M$ 이고 X_j 는 Z_k 를 크기 순으로 정렬한 값이다. 그러므로 검정통계량은 식 (4.31)과 같이 주어지며 식 (4.32)로 정의되는 신뢰 한계 안에 있게 되면 유의수준 α 에서 경향을 갖는다고 할 수 있다.

$$S = X_{\omega(M+1)/2} \quad \text{if } \omega M \text{ is odd} \quad (4.31a)$$

$$S = \frac{1}{2}[X_{\omega M/2} + X_{(\omega M+2)/2}] \quad \text{if } \omega M \text{ is even} \quad (4.31b)$$

$$[X_{M_L}, X_{M_U}] \quad (4.32)$$

여기서

$$M_L = \frac{\omega M - u_{1-\alpha/2} [V(SS)]^{1/2}}{2} \quad (4.33a)$$

$$M_U = 1 + \frac{\omega M + u_{1-\alpha/2} [V(SS)]^{1/2}}{2} \quad (4.33b)$$

[예제 4.9] 대관령 자료에 대한 seasonal Kendall test 결과는 아래와 같다.

식 (4.26a)와 (4.26b)에서 $SS=231$, $V(SS)=8364$ 이므로 상향 경향을 갖는다. 식 (4.27)로부터 검정통계량 $U_c=2.526$ 이므로 유의수준 $\alpha=0.05$ 에서 경향을 갖는다고 할 수 있다. 또한 $\omega=12$, $M=153$ 이므로 식 (4.31)로부터 $S=1.711$ 이며, 식 (4.33)의 $M_L=828$, $M_U=1007$ 을 이용하여 구한 하한계와 상한계는 각각 $X_{M_L}=0.315$, $X_{M_U}=3.189$ 이므로 대관령 월평균강수량은 경향을 갖는다고 할 수 있다.

• 분산에 대한 경향 분석

주어진 계절자료($Y_{\nu, \tau}$, $\nu=1, \dots, N$; $\tau=1, \dots, \omega$)의 평균뿐만 아니라 분산도 경향을 가질 수 있으며 이를 분석하기 위하여 다음과 같이 변수를 정의한다.

$$Z_{\nu, \tau} = Y_{\nu, \tau} - a - b_1 i + b_2 i^2 - \dots - b_2 i^H \quad (4.34)$$

식 (4.34)에서 새롭게 정의된 변수 $Z_{\nu, \tau}$ 는 원자료에서 평균에 대한 경향을 제거한 변수임을 알 수 있다. 이렇게 평균에 대한 경향을 제거한 자료를 이용하여

$$V_\nu = \sqrt{\frac{1}{\omega-1} \sum_{\tau=1}^{\omega} (Z_{\nu, \tau} - \hat{\mu}_\tau)^2} \quad (4.35)$$

$\nu=1, \dots, N$ and $\tau=1, \dots, \omega$

$$U_\nu = \nu. \quad (4.36)$$

여기서

$$\hat{\mu}_\tau = \frac{1}{\omega} \sum_{\nu=1}^{\omega} Z_{\nu, \tau} \quad (4.37)$$

를 정의하면 연평균 표준편차 V_ν 는 다음과 같은 다항식으로 표기할 수 있다.

$$V_\nu = c_0 + c_1 U_\nu + c_2 U_\nu^2 + \dots + c_G U_\nu^G \quad (4.38)$$

여기서 c_0, \dots, c_G 는 계수이며, $G=1$ 이면 V_ν 가 선형적 경향, $G>1$ 이면 비선형 경향을 갖는다.

[예제 4.10] 대관령 월평균강수량자료의 표준편차의 경향성 분석 결과는 아래와 같다.

앞에서 설명한 절차를 이용하여 차수 $G=1$ 로 나타났으며, 계수는 각각 $c_0=114.21$, $c_1=4.155$ 이었다. 그러므로 월평균강수량자료의 표준편차는 다음 식과 같은 선형 경향을 갖는다고 할 수 있다.

$$V_\nu = 114.21 + 4.155 \nu, \quad \nu=1, 2, \dots, 18$$

4.3 맺음말

본 강좌에서는 수문자료의 평균과 분산값의 변동 및 경향을 분석하는 방법에 대하여 간단히 설명하였다. 여기서 소개된 방법은 매개변수적 방법과 비매개변수적 방법으로 분류될 수 있으며, 매개변수적 방법을 사용할 때의 기본 가정은 주어진 자료가

정규분포라는 것이다. 비정규분포를 갖는 자료에 매개변수적 방법을 이용하여 변동 및 경향을 분석하는 것은 동일한 자료에 비매개변수적 방법을 적용하는 경우보다 정확도가 낮은 것으로 알려져 있다(Bradley, 1968). 자료가 비정규분포인 경우에는 정규분포로 변환(transformation)한 후 매개변수적 방법을 적용하는 것이 바람직하며, 만약 원자료를 그대로 사용하는 경우에는 비매개변수적 방법을 사용하는 것이 더 신뢰할 수 있는 결과를 얻을 수 있다.

다음 강좌에서는 우리가 일반적으로 수문자료에서 자주 보게 되는 결측치를 보완하는 방법 및 자료확충 방법에 대해서 알아보기로 한다.

참 고 문 헌

- Bradley, J.V. (1968). *Distribution-Free Statistical Tests*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*, John Wiley & Sons, New York.
- Gilbert, R.O. (1987). *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold Co., New York.
- Helsel, D.R. and Hirsch, R.M. (1992). *Statistical Methods in Water Resources*, Elsevier.
- Hirsch, R.M., Slack, J.R., and Smith R.A. (1982). "Techniques of trend analysis for monthly water quality data." *Water Resour. Res.*, Vol. 18, No. 1, pp. 107-121.
- Hirsch, R.M., Helsel, D.R., Cohn, T.A., and Gilroy, E.J. (1993). *Statistical analysis of hydrologic data in Handbook of Hydrology* ed. Maidment, McGraw Hill, New York.
- Hollander, M. and Wolfe, D.A. (1973). *Nonparametric Statistical Methods*, Wiley, New York.
- Mann, H.B. and Whitney, D.R. (1947). "On a test of whether one of two random variables is stochastically larger than the other." *Annals of Math. Statistics*, Vol. 18, No. 1, pp. 50-60.
- McLeod, A.I., Hipel, K.W., Camacho, F. (1983). "Trend assesment of water quality time series." *Water Resour. Bull.*, Vol. 19, No. 4, pp. 537-547.
- Sen, P.K. (1968). "Estimates of the regression coefficient based on Kendall's tau." *J. of American Statistical Association*, Vol. 63, pp. 1379-1389.
- Tukey, J.W. (1977). *Exploratory Data Analysis*, Addison-Wesley, Readings, Mass.
- Wilcoxon, F. (1945). "Individual comparisons by ranking methods." *Biometrics Bulletin*, Vol. 1, No. 6, pp. 80-83. ♣