

우리나라 강수량과 기온자료의 통계적 특성과 기후변화의

통계적 증거에 관한 연구

이 상 훈 (수원대)

장 영 기 (수원대)

1. 서론

최근에 대기오염과 관련하여 기후변화가 자주 논의되며 기후변화에 대한 논문도 다수 발표되고 있다. 이들 결과들을 요약하면 연평균기온은 상승이 뚜렷하며 연강수량은 변동이 심해서 결론을 내릴 수 없다고 보고되고 있다 (중앙기상대, 1989). 기후변화에 관한 연구들이 사용한 연구방법과 자료의 기간은 연구자마다 다르며 아직 합의된 기준이 제시되고 있지 못하다.

연구방법들을 개괄해 볼 때에 세가지 유형으로 분류할 수 있다. 첫째는 두 기간의 평균값을 수치로 비교하는 방법이다. 이 때의 기간은 10년, 20년, 30년등으로 연구마다 다르다. 두번째 방법은 일정기간의 이동평균치를 그래프에 찍어서 시각적으로 변화를 보이는 방법이다. 이때 이동평균은 대개 5년평균값을 구한다. 세째는 일정기간의 측정치들을 선형회귀분석하여 직선의 기울기를 구하고 그 기울기를 년변화율로 해석하는 방법이다. 이들 방법들은 각각 다음과 같은 문제점이 있다. 첫째는 자료의 기간의 문제이다. 평균치는 자료의 수에 따라 달라지며, 이동평균도 3년 또는 7년으로 한다면 달라질 것이므로 최소한의 기간에 대한 기준설정이 요구된다. 둘째는 어느 정도 차이가 나와 유의미한 차이라고 말할 수 있는가 하는 문제이다. 평균기온의 차이가 0.001°C 이라면 차이가 있다고 보아야 할 것인가 없다고 보아야 할 것인가? 회귀분석식의 기울기가 0.001 이라면 증가추세라고 말할 수 있는가? 또는 차이가 없다고 말해야 하는가? 이 연구에서는 기후변화의 증거를 제시하기 위하여 지금까지 많이 보고된 통계기법으로서 모수적검사(parametric test) 외에도 비모수적검사(non-parametric test)들을 조사하여 기후변화의 증거로 사용할 수 있는 기준을 제시한다.

2. 이론적 고찰

두 기간의 평균치의 차이가 통계적으로 의의있지는 t-test로써 판단할 수 있다.

회귀분석은 한 기간의 측정치에 대해 연도별 변화의 크기를 회귀식의 기울기로써 계산할 수 있다. 연구자들이 이들 모수적 통계기법을 이용할 경우 자료의 통계적 특성을 기본전제에 비추어 확인해야한다. t-test의 경우 두표본은 독립적이어야하며 측정치들은 정규분포를 이루어야 한다는 가정이 요구된다. 회귀분석의 경우 측정치들은 정규분포를 이루어야 하며, 시계열로 놓고 볼 때에 앞뒤의 측정치는 서로 독립적이어야 한다. 여기에서 공통적으로 중요한 것은 표본이 정규분포를 이루어야 한다는 가정이다.

비모수적검사는 표본의 분포에 상관없이 적용할 수 있는 통계검사로써 환경연구 분야에서는 많이 사용되고 있지 않으나 다음과 같은 경우에는 모수적검사보다 장점이 있다. 1)표본의 정규성이 의심스러울 때, 2)표본의 크기가 작을 때, 3)극치(outlier)가 있을 때, 4)계절성(seasonality)이 있을 때. 비모수적검사는 측정치를 크기 순서로 늘어 놓은 후 정해지는 순위(rank)를 새로운 변수로 만드는 방법, 두 측정치의 차이값의 부호(sign)만을 세어서 +부호의 수와 -부호의 수를 새로운 변수로 만드는 방법 등이 있는데 기후변화연구에 적용할 수 있는 검사로서 다음 세가지를 소개한다.

2.1 Mann-Whitney test

이 검사는 모수적검사인 t-test에 상응하는 비모수적 검사로서 두 표본의 순위합(rank sum)의 차이가 통계적으로 의의가 있는가를 검사한다. 먼저 두 표본의 측정치 모두를 한 집단으로 보고 각 측정치의 크기에 따라 제일 작은 값을 1, 그 다음 작은 값을 2, . . . 이러한 식으로 순위를 매긴다. 변수 X(표본의 크기 n)와 Y(표본의 크기 m)의 순위를 각각 $R(X_i)$, $R(Y_i)$ 라고 하면 Mann Whitney 검사치는 다음과 같이 계산된다.

$9 \leq n \leq 20$ 일때

$$U = n*m + (n*(n+1))/2 - R_1 \quad \text{여기서, } R_1 = \sum R(X_i) \quad \text{————(1)}$$

$n > 20$ 일때

$$Z = (U - n*m/2) / \sqrt{\{n*m*(n+m+1)\}/12} \quad \text{———— (2)}$$

이렇게 계산된 검사치는 (1)식의 경우 표본의 크기와 신뢰수준에 따라 달라지는 기준치(critical value)를 제시한 표의 값과 비교되며 (2)식의 경우 Z값은 표준화된

Z점수로서 해석된다.

2.2 Mann's test

회귀분석식에서 얻어지는 계수 b 값은 두가지 정보를 제공한다. 변화의 방향, 즉 시간에 따라 측정치가 증가하는 경향인지 아니면 감소하는 경향인지는 b 의 부호에 의해 결정되며, 변화의 크기는 b 의 절대값의 크기이다. 비모수검사로서 Mann's test는 변화의 방향을 검사하는데 측정치가 무작위로 발생한다는 영가설 (null hypothesis)을 검증한다.

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(X_j - X_k) \quad \text{--- (3)}$$

$$\text{여기서, } \text{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases}$$

만일 X 가 변화의 경향성이 없이 완전히 무작위로 발생한다면 S 값은 0이 될 것이며 0에서 멀어질수록 X 는 무작위가 아니고 어떠한 경향성을 나타낸다고 말할 수 있다. 사례 수 n 이 10을 넘으면 S 는 정규분포에 접근하며 Z 값과 같은 모수적해석을 적용할 수 있다. 이 검사는 단지 변화의 경향성만을 확인할 뿐 변화의 크기는 알 수 없다는 단점을 가지고 있다.

2.3 Median Slope Change Estimator (MSCE)

변화의 크기를 나타낼 수 있는 비모수검사로서 MSCE를 제안한다. MSCE는 Hirsch 등(1982)에 의해서 제안된 월별자료에 적용할 수 있는 검사를 연별자료에 적용한 것으로서 다음과 같이 정의된다.

모든 (X_j, X_k) 쌍에 대해,

$$d_{jk} = (X_j - X_k)/(j-k) \quad \text{여기서, } 1 \leq k < j \leq n \quad \text{--- (4)}$$

MSCE = d_{jk} 값들의 중앙값

자료수가 n 일 때 모두 $\sum_{i=1}^{n-1} i$ 개의 d_{jk} 값이 계산되며 MSCE는 X 값의 연변화율로 해석할 수 있다.

3. 자료와 연구방법

서울, 부산, 인천, 목포등 4개 지점의 연평균기온과 연강우량의 83년간 자료 (1908-1990)를 수집하여 SAS의 PROC UNIVARIATE, PROC TTEST, 그리고 PROC REG를 이용하여 분석하였으며 비모수검사는 BASIC Program을 작성하여 분석하였다.

4. 결과와 토의

표1과 표2를 보면 연평균기온은 정규분포를 가정할 수 있으며 연강우량은 비정규 분포임을 알 수 있다. 연평균기온에 대한 t-test와 회귀분석결과는 표3, 표4와 같다.

표 1. 연평균기온의 통계적 특성 (1910-1990)

지역	n	평균	표준편차	왜곡도	변동계수	D	Prob>D	ACF(1)	ACF(2)	ACF(3)
서울	81	11.295	0.730	-0.086	6.5	0.983	0.750	0.391	0.319	0.137
인천	81	11.097	0.619	-0.513	5.6	0.977	0.435	0.152	0.190	-0.004
부산	81	13.774	0.589	-0.066	4.3	0.982	0.663	0.283	0.283	0.065
목포	81	13.413	0.519	0.089	3.9	0.980	0.562	0.177	0.171	-0.066

표 2. 연강우량의 통계적 특성 (1908-1990)

지역	n	평균	표준편차	왜곡도	변동계수	D	Prob>D	ACF(1)	ACF(2)	ACF(3)
서울	83	1292.5	333.6	0.614	25.8	0.965	0.100	0.024	-0.019	-0.040
인천	83	1098.1	278.7	0.837	25.4	0.951	0.009	0.014	-0.064	-0.011
부산	83	1424.4	345.0	0.381	24.2	0.959	0.034	-0.172	0.007	-0.219
목포	83	1104.9	265.2	0.384	24.0	0.953	0.014	-0.122	0.005	-0.105

주: D는 Kolomogorov D statistic 을 의미하며 분포의 정규성을 검사한다.

Prob>D 는 분포가 정규분포일 확률을 나타낸다.

ACF(L)는 지체 L에서의 자기상관계수를 나타낸다.

표 3. 연평균기온에 대한 t-test 결과 (1920-1949, 1961-1990)

지역	평균1	평균2	차이	t	Prob> t	Significance
서울	10.975	11.814	0.839	5.706	0.0001	Yes
인천	10.973	11.376	0.403	2.806	0.0068	Yes
부산	13.583	14.101	0.518	4.121	0.0001	Yes
목포	13.293	13.606	0.313	2.571	0.0127	Yes

표 4. 연평균기온에 대한 선형회귀분석 결과

지역	n=81 (1910-1990)				n=30 (1961-1990)			
	b	T	Prob	d	b	T	Prob	d
서울	0.0192	7.01	0.0001	1.87	0.0210	1.78	0.0867	1.74
인천	0.0125	4.82	0.0001	2.11	0.0158	1.41	0.1703	1.66
부산	0.0139	5.95	0.0001	1.90	0.0207	1.89	0.0691	1.69
목포	0.0088	3.89	0.0002	1.84	0.0047	0.43	0.6704	1.50

주: T는 영가설 $H_0: b=0$ 에 대한 값, b는 회귀식의 기울기이다.

Prob value는 영가설을 받아들일 수 있는 확률이다.

연강수량은 정규분포를 가정할 수 없으므로 t-test나 회귀분석을 할 수 없다. 대신 두 기간 (1920-1949, 1961-1990)에 대해 비모수적검사를 적용한 결과는 표5와 같으며 평균의 차이를 발견할 수 없었다.

표 5. 두 기간의 연강수량에 대한 Mann-Whitney test 결과

지역	R_1	U	Z	Prob>Z
서울	848.5	516.5	0.983	0.164
인천	848	517	0.990	0.161
부산	868	497	0.695	0.244
목포	901	461	0.162	0.436

연평균기온자료에 대하여 회귀분석과 MSCE를 함께 적용하여 분석할 결과는 표7과 같다.

표 7. 연평균기온에 대한 회귀분석과 MSCE의 비교 (1961-1990)

지역	서울	인천	부산	목포
b	0.021	0.016	0.021	0.005
MSCE	0.024	0.017	0.020	0.004

두 분석방법은 비슷한 결과를 보여주는데 이것은 비모수적검사인 MSCE가 회귀분석에 비해 그 적용범위가 더 넓으며 그 효율성에 있어서도 우수하다고 해석된다. 앞으로 기후변화의 연구에 있어서 적용할 수 있는 통계적 준거로서 활용할 수 있을 것으로 생각된다.

참 고 문 헌

- 중앙기상대 (1989), 지구기후를 진단한다, 206 pp.
- 김 광식 (1988), 도시화가 강수현상에 미치는 영향, 대한토목학회지 36(2), 3-7.
- 김 승, 김규호 (1989), 한국년평균강수량의 추정, 수공학논총 31, 5-16.
- 이 희연 (1989), 지리통계학, 범문사, 763 pp.
- Conover, W. J. (1971) Practical Nonparametric Statistics. John Wiley & Sons Inc. pp. 462.
- Hirsch, R. M., J. R. Slack, and R. A. Smith (1982) Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research, 18(1). pp. 107-121.
- Mann, H. B. (1945) Non-parametric Tests Against Trend. Biometrika, 13. pp. 245-259.
- Wonnacott, T. H. and R. J. Wonnacott (1977) Introductory Statistics. John Wiley and Sons. 650 pp.