

A Study of Statistical Methods in the Water Environmental Research of Han and Nakdong River Basins¹⁾

SangBock Lee²⁾ · Malsuk Kim³⁾

Abstract

This paper provides the checklist of statistical methods in the water environmental research of Han and Nakdong river basins in South Korea. There are many errors pointed out in adopting statistical methods for the researches, as an example, basic statistical assumptions are missed for t-tests or regression analyses. Some new ideas are proposed for better researches of the river basins in Korea.

Keywords : checklist, Han and Nakdong river, water environment

1. 서론

한강, 낙동강 광역수계유역의 환경분야 연구는 물 문제가 중심연구로 최소한의 수질환경기준 2 등급을 만족시키기 위한 다양한 수질환경 연구들이 수행되고 있다. 현실적으로 수도권, 지방자치단체간 또는 산업간 경쟁적인 개발정책으로 인하여 수자원 이용은 물론 댐 건설, 수질오염 등을 비롯한 수많은 갈등이 끊임없이 야기되고 있다. 최근 10여 년 동안 미해결 문제로 남아 있는 영남갈등의 한 사례로 대구광역시의 위천공단조성 문제, 신규 댐 건설에 따른 민원의 발생은 지역간의 갈등과 수질환경분쟁의 구체적인 사례이다.

남한의 4대 강 광역수계에 대한 수량 및 수질 등의 수자원 환경의 조사와 연구는 국가기관, 지방자치단체, 대학, 기업체산하 각종 연구소별로 이루어지고 있으며, 국가

1) This research was supported by the Catholic University of Daegu Research Grants in 2001

2) Professor, Faculty of Information and Environmental Science, Catholic University of Daegu

3) Lecturer ,Institute of Continued Education, Catholic University of Daegu

기관과 지방자치단체는 주로 기초통계생산과 수자원 정책입안을 하고 있으며, 대학과 연구소에서는 수량 혹은 수질에 관한 예측모형, 수자원 평가방법과 수자원 정책연구가 주류를 이루고 있다. 현재 한강과 낙동강 유역의 수질측정과 분석 및 결과공표는 환경정책기본법 15조, 수질환경보존법 제 3조, 상수원 관리규칙 제 23 조의 2에 따른다. 한강의 경우, 한강유역환경관리청을 비롯하여 관할 4개 기관에서 관리하는 181개 측정지점의 자료에 대한 데이터베이스가 구축되어 한강유역 환경관리청 홈페이지 상에 다양한 기술통계를 제공하고 있다. 낙동강유역의 경우에 285개 측정지점이 운영되고 있다. 이러한 축적된 방대한 수질측정자료에 대하여 다변량통계분석과 시계열분석 방법 등을 활용한 연구들도 적으나 몇 편 있다. 특히, 광역수계환경 예측모형 연구는 유역의 개발정도, 산업형태, 인구 밀도, 지형 및 지질, 수질관리정도 등의 다양한 하천유역의 특성과 계절별 하천수량의 변화 그리고 수질측정과 채수 표본지점 설정이라는 다양한 통제변인과 환경요인이 가정되어 있으며, 이러한 변인들을 고려한 표본조사, 실험계획법, 다변량분석 등의 고도의 통계분석기법을 필요로 한다. 박배경과 박석순(1995)은 하천수질과 유역이용도간의 집단 자료의 비교는 일반적 단순통계분석으로 비교할 수 없으므로 함수관계를 이용한 중회귀분석을 시도하였다. 그밖에 수치모델로 미국에서 개발된 QUAL2, WASP 혹은 생태계 예측모델을 사용한 논문들도 다소 있다. 그러나 국내 기후 특징과 환경청이나 연구소의 축적자료를 고려한 수질환경 분석방법과 수자원 및 수질예측모형연구는 아직 미미한 실정이다.

본 연구에서는 1992년부터 2001년까지 과거 10여 년간에 국내의 대표 환경연구학술지인 한국 물(수자원)환경학회의 수질보전과 한국환경과학회지에 수록된 한강, 낙동강 광역수계 환경연구에 사용된 통계분석 기법을 점검표를 활용하여 정리요약하고, 이 자료를 통해 나타난 연구가설의 타당성 평가, 통계분석기법의 활용과 통계분석기법의 오용사례를 정리하며, 나아가 광역하천수계 환경자료에 대하여 보다 객관적인 자료분석과 수질예측모형을 위한 일반적 통계분석기법 활용의 개선방안을 제시하고자 한다.

2. 점검표

2.1 점검표의 구성

점검표(checklist)는 통계조사를 실시하고 통계분석을 하는 기관 및 산업체 혹은 통계적 방법을 활용하는 연구자가 통계학에 대한 이해의 부족에서 범할 수 있는 통계적 오류를 가능한 줄일 수 있도록 고려해야 할 사항을 참조하여 통계자료조사 및 통계분석기법을 보다 효율적으로 활용할 수 있도록 한 도구이다.

<표 2-1> 1단계 - 연구설계

조사설계	정확하게 수행됨	잘못수행됨	해당사항없음	실험설계	정확하게수행됨	잘못수행됨	해당사항없음
	O	X	N		O	X	N
1. 연구문제와 연구가설 2. 연구대상 모집단과 조사집단 조사집단의규모,구조,특성 3. 표본설계 표본추출방법 표본의 크기 표본의 대표성	O	X	N	1. 연구문제와 연구가설 2. 실험설계 실험계획설계 실험의 통제 반복화 확률화 블럭화	O	X	N

한강, 낙동강 광역수계환경 학술연구논문에 대하여 이상복과 김말숙(1998)의 타당성 평가, 화이트의 통계적 평가법(1979), 윤기중외 2인의 점검표(1987), 최종후와 이재창(1990)의 점검표를 기초로 하여 3단계 점검표를 사용하였다. 1단계 연구설계는 <표 2-1>과 같이 연구방법에 따라 조사설계와 실험설계로 분리하여 제시한다. 2단계에서는 통계적 기법을 이용하여 결론을 도출하기까지의 전반적인 점검 항목으로 구성되어 통계분석 기법의 타당성 평가와 결과해석을 위한 것으로 <표2-2>에 정리되어 있다. 점검항목별 판정은 ‘정확하게 수행됨’, ‘잘못 수행됨’, ‘해당사항 없음’ 3개 범주로 한다.

<표2-2> 2단계 - 통계분석과 결과 해석

점검항목	정확하게수행됨	잘못수행됨	수	해당사항없음
1. 통계분석기법의 선택 1) 주용 통계량 2) 분석절차 3) 계산 프로그램(패키지) 2. 결과해석 1) 연구가설과 통계적 결론의 연관성 2) 통계정보 그래프스	O	X		N

2. 2 점검항목의 일반적 제 문제

모든 연구의 첫 단계는 연구문제(research problem) 또는 주제를 발견·선정해서 제기하는 작업이 된다. 특히 수질환경에 관한 연구문제는 이론적 체계와 관련시켜 설정되어야 하므로, 연구문제를 서술할 때에는 반드시 연구문제와 이론적 체계와의 관계를 서술하여야 되며 그 문제에 관한 기존의 연구도 제시해야 된다. 즉 연구문제가 어떤 맥락에서 나오게 되었고 왜 그러한 연구가 필요한지가 설명되어야 한다. 이와 함께 연구가설은 가설로서의 요건을 갖추어야 하며 연구문제에 대한 가장 적합한 해답이 될 수 있어야 하고, 이론적 체계와 관련된 간단 명료하면서도 주어진 자연적 혹은 인위적 환경변인들간의 관계가 명시되어야 하는 동시에 그 변인들은 명확하고 측정 가능하여야 하며, 실증적으로 검증가능(verifiable or testable)한 것이어야 하며, 주어진 연구의 방향을 결정한다. 따라서 적절하고 타당한 가설을 세우는 것은 모든 연구의 핵심이다.

실무적인 어려움 때문에 정의된 모집단과는 약간 다른 연구(조사)모집단(study population)을 정의할 때 연구모집단의 원소, 표본추출단위(sampling unit), 범위(extent:지역, 세계), 시간 등의 관점에서 명백하고도 한정적으로 정의되어야만 한다.

표본추출방법이 결정되면 수질환경연구에 있어 실험체수의 추출방법과 비용에 따라 표본의 크기를 결정하여야 하며 이때 표본의 크기, 오차한계, 신뢰계수에 대한 언급이 필요하다. 체수표본추출인 경우에는 표본은 연구목적과 방법에 의해서 결정되는 모집단을 대표할 수 있도록 추출되어야 하며, 표본설계에서 표본은 모집단을 대표할 수 있다는 것이 인식되도록 설명되어야 한다.

실험연구에서의 연구문제는 대체로 그 성질에 따라 두 가지로 나누어 볼 수 있는데 하나는 아직 연구된 적이 없고 연구자의 호기심에서 출발되는 탐색적 실험연구(exploratory research)를 위한 문제요, 또 하나는 이론적으로 예측된 가설이나 또는 다른 연구에서 발견된 사실을 실증적으로 검증해 보고자 하는 검증적 실험연구(confirmatory experiment)를 위한 문제이다. 탐색적 실험연구의 문제들은 새로운 변인들간의 관계를 알아보려는데 목적이 있는 것이라면 검증적 실험연구의 문제들은 이미 알려진 변인들간의 관계를 확증해 보려는데 그 목적이 있다. 수질연구의 실험설계에서도 주어진 연구문제와 연구가설을 검증하기 위한 구체적인 실험방안을 선정하고, 그에 따라 환경적 혹은 시간적인 가외변인을 통제하면서 실험처치를 가하고, 검사 또는 측정을 통하여 자료를 수집하고, 그것을 분석하는데 필요한 모든 방법과 절차들을 결정한다. 또한 반복(replication), 확률화(randomization), 블록화(blocking)는 통계적 실험에서 반드시 고려되는 기본원리이다.

실제적인 데이터 분석은 자료의 검정이나 추정 같은 전통적인 통계적 추론인 확증적 데이터분석(CDA: Confirmatory Data Analysis)의 단계와 데이터의 구조와 특징을 알아내기 위한 탐색적 데이터분석(EDA: Exploratory Data Analysis)의 단계로 구분된다. 먼저 수집된 자료를 탐색적 자료분석을 통하여 데이터의 구조와 특징을 파악을 목적으로 이를 효과적이고 신뢰성 있는 데이터의 요약과 정보통계그래픽스 기법이 사용된다. 또한 데이터가 간직하고 있는 정보의 손실 없이 데이터의

특성을 더욱 쉽게 파악할 수 있도록 데이터를 수량화 등의 변환(transformation)하는 것이 때때로 유용하다. 그러므로 연구자는 데이터를 일차적으로 요약하고 난 후 추정 및 가설검정에 들어가기 전에 데이터를 면밀히 살펴보는 과정이 필요하다.

타당성 있는 통계적 기법을 적용하여 얻어진 분석결과의 해석은 주어진 연구문제에 논리적인 추론을 하고 결론을 도출하는 것은 아주 중요하고 또한 어려운 일이다. 주요 통계량에 대한 설명과 유의수준을 기술하고 정보통계그래픽스를 활용한 보조적인 해석을 함께 검토할 필요성도 있다.

3. 한강, 낙동강 광역수계환경 학술연구에 적용된 통계분석기법의 활용 사례

한강, 낙동강 광역수계유역의 환경학술연구분야의 통계적기법의 활용도와 타당성을 평가하기 위하여 환경연구학술지인 한국 물환경학회의 수질보전, 한국환경과학회지, 한국환경분석학회지 3종을 대상으로 조사하였다. 학술지 선정은 학술진흥재단 국내 전국규모 등재후보 학술지이거나 또는 통계적 기법의 이용빈도가 높은 학술지를 대상으로 선정하였다. 1992년부터 2001년까지 10년간 발간된 학술지를 조사한 결과이며 한국환경분석학회지는 1998년부터 발간되었기에 그 후의 자료를 조사하였다.

관련논문의 수가 적은 한국 환경분석학회지의 경우에는 점검표 분석에서 제외하여 연구의 참고자료로만 활용하였다. <표3-1>에서는 전체 게재논문 1217편 가운데 한강, 낙동강 광역 수계에 관련된 논문은 60 편으로 4.9% 정도이다. 그리고 다수의 연구자들이 공동연구(평균 3.6명/편)를 한 사실로 미루어 연구완성도가 높은 논문으로 볼 수 있다. <표3-2> 수질보전의 경우, 기초통계 사용빈도는 9.28회, <표3-3> 한국환경과학회의 경우는 논문편당 7.57회로 모든 학술논문들이 기술통계분석을 많이 사용하고 있음을 알 수 있다. 그러나 대부분 자료의 요약, 정리 수준인 기술통계에 머물고 있으며, 추론통계를 적용한 논문 편수는 15편으로

<표 3-1> 학술지별 한강, 낙동강 광역수계에 대한 학술연구논문비율(1992년-2001년)

학술지	광역수계 관련논문비율	관련논문 공저자수비율	비고
수질보전	32/491(6.5%)	109/32(3.4명/편)	현 년 4회 발간 현 년 6회 발간 현 년 4회 발간 (1998년-2001년자료)
한국환경과학회지	28/726(3.9%)	107/28(3.8명/편)	
한국환경분석학회지	4/149(2.7%)	17/4(4.3명/편)	

비율은 관련논문수 / 총게재논문수로 표현하였다.

<표 3-2> 수질보전의 통계적 기법과 정보통계그래픽스 활용도

조사 논문수	이용통계기법				정보통계그래픽스			
	기술통계	시계열	추론통계	기타	깍은선 그래프	막대 그래프	산점도	기타
32	212/32	38/11	31/13	14/9	60/14	27/10	15/6	9/3

해당항목 비율은 누적도수 / 논문수로 표현하였음. 기타는 요인분석 1회, 수리모형은 QUEL2E, HEC-2, WASP 사용, 정보통계그래픽스의 기타는 통계지도임.

전체 논문의 25% 정도이다. 다변량 분산분석, 중회귀분석, 인자분석, 판별분석 등 고급통계분석기법을 사용한 연구는 10여 편 미만의 몇몇 연구자로 국한되어 있어 환경자료에 대한 통계기법 활용 면과 환경통계교육에 많은 문제점이 있음을 알 수 있었다.

정보통계정보그래픽스 사용에 있어 깍은선그래프는 총 29편으로 전체 논문(60편)의 50%정도를 차지하여 편당 8.04회, 산점도 21편(30%)으로 편당 3.1회, 막대 그래프의 경우 14편(23%)으로 편당 3.79회 정도 사용되고 있어, 광역수질환경자료 논문의 경우에 그래픽스를 통한 연구결과분석이 상당히 많이 있음을 알 수 있다.

<표3-4>의 분석기법별 활용도를 보면 평균, 비율 등 대표값의 활용도는 높으나 분산, 표준편차, 혹은 범위를 나타내는 산포도의 활용도는 대표값 활용도의 20% 정도 수준으로 아주 낮은 편이며 추론 통계기법에서는 회귀분석을 가장 많이 사용하고 있다. 계절별 큰 차를 보이는 수질환경자료의 특징상 시계열분석 기법은 <표3-2, 3>에 나타내었다. 수질보전의 경우 11편(39%), 한국 환경과학회지의 경우는 10편(35.7%)로 시계열자료분석이 차지하고 있었다.

<표 3-3> 한국환경과학회지의 통계적 기법과 정보통계그래픽스 활용도

조사 논문수	이용통계기법				정보통계그래픽스			
	기술통계	시계열	추론통계	기타	깍은선 그래프	산점도	막대 그래프	기타
28	297/28	57/10	84/12	17/7	109/15	38/11	35/8	1/1

해당항목 비율은 누적도수 / 논문수로 표현하였음. 기타는 군집분석(2),인자분석(1), 로지스틱(1), 수리모형으로 QUAL2E, HEC-2, WASP, 생태계 예측모형을 사용, 통계그래픽스의 기타는 통계지도임.

<표3-4> 통계분석 기법별 활용

학회지	논문수	대표값	산포도	도수 분포표	회귀 분석	상관 분석	t검정	신뢰 구간	기타
수질보전	32	99	38	6	28	7	5	27	분산분석(2) 수리모형(10)
한국환경과학 회지	28	101	4	.	10	8	7	2	분산분석(3) 군집분석(1) 수리모형(1)
계	60	200	42	6	38	15	12	29	(17)

4. 한강, 낙동강 광역수계환경 학술연구에 적용된 통계분석기법의 타당성 평가

본 절에서는 수질보전지와 한국환경과학회지의 1992년에서 2001년 사이에 수록된 한강, 낙동강 광역수계연구 논문 가운데, 통계적 기법을 활용한 32편, 28편의 논문에 대하여 각각 수1), 수2) 또는 환1), 환2) 등으로 임의로 번호를 표기하였다.

4. 1 수질보전의 통계적 기법 활용의 타당성 평가

가. 1 단계(연구설계)

1) 연구문제와 연구가설

연구문제의 서술은 조사설계 및 실험설계의 모든 학술논문에 있어 대체로 잘 정리되어 있었다.

분석결과	판정	O	X	N	소계
	논문편수	32	.	.	32
	%	100	.	.	100

본 논문에서 배열한 수19)의 연구가설의 가정은 “...자료는 항상 일정하지 않으나 이후의 해석은 변화가 없는 안정상태를 가정...”로 연구가설의 가정을 잘 기술한 경우로 대다수의 논문에서 이 정도의 가정을 서술하고 있다. 수13)의 경우 연구가설이 불일치인 경우“... 수돗물보다는 어류 등 수중동물을 식용하였을 경우에 더 크다...”로 이 논문의 경우 하천의 수질오염에 대한 연구로 수돗물 혹은 수중동물에 의한 중독연구와는 연구가설에서 제시된 바 없었다.

2) 연구의 대상

모집단과 조사집단, 그 크기 규모, 특징을 정의하는 것은 연구의 범위를 명백히 하는 문제로 매우 중요하다. 대부분의 논문에서는 해당 지역의 지도도표를 첨부하여

잘 기술하고 있다. 한 예로 수1)에 있어 “...대상하천으로는 유역의 토지이용 상황이 농촌, 도시, 복합지역으로 구분 가능하다고 판단되는 중랑천, 진위천 및 한강 하류부를 선택하 ...” 처럼 표본모집단을 언급하였다.

분석결과	판정	O	X	N	소계
	논문편수	31	1	.	32
	%	97	3	.	100

3) 표본설계

표본추출방법이 제대로 기술된 예로 수2)의 “...조사기간은 4계절을 고려하여, ...5월, 8월, 9월 및 11월에 1회씩 강우기를 피하여 하천의 중심부에서 표층수를 채취...”와 표본지도의 사례로 “...호수의 전면에 걸쳐 대표적인 8지점을 선정하여 조사지점 Fig.1과 같다...”로 대부분 잘 기술하고 있으나, 수4)의 경우 “...계절적 영향을 포함한 대표성을 부여하기 위한 작업이 어려워 조사가 편리한 세대를 대상으로...” 표본의 대표성을 잃어버린 경우이다.

분석결과	판정	O	X	N	소계
	논문편수	30	2	.	32
	%	94	6	.	100

4) 실험설계

수질연구의 실험의 경우 환경법의 실험규정에 따라 반복, 확률화, 블록화는 대체로 잘 지켜지고 있었다. 사례로 수2)의 “...환경오염공정시험법(수질분야,수질편)에 준하여 측정...” 표현과 수24)의 경우 요인별 수준을 잘 정리한 사례로 “...산업용화학(C-1,2,3), 고무 및 플라스틱(R-1,2),...” 들 수 있다. 그러나, 수15)의 경우 1988, 1990, 1992, 1993, 1994년도의 측정자료의 측정시점과 측정회수가 불일치함으로 단정적인 비교를 내리기 어렵게 한다. 수23)의 경우에, 단순 실험결과인 평균값과 비율만으로 비교하여 단정하는 것은 일반화를 어렵게 하는 원인으로 실험계획법에 따른 분산분석 등의 기법이 필요하다.

분석결과	판정	O	X	N	소계
	논문편수	31	1	.	32
	%	97	3	.	100

나. 2 단계(분석 및 결과해석)

1) 통계분석기법의 선택과 정보통계그래픽스의 사용

대체로 기술통계량을 많이 사용하여 단순 통계분석의 수준에 머무르고 있으며, 간단한 도표 자료의 경향을 보고 추론을 하려는 경우도 있었다. 수량화의 사례로 수11)의 “...각 소유의 면적의 비를 설명변수... 하천수질을 반응변수로...” 적절한 수량화를 한 경우로 보이며, 사용된 전산프로그램의 소개로 수5)의 경우에 “...QUAL II모델은... 유한차분법을...” 등 비교적 모델원리를 자세하게 소개하고 있다. 수6)의 S

negative regression 기법으로 S/u의 자료분석을 하였으나 결론과 적절한 이에 대한 설명이 없는 부적절한 분석 방법으로 보여진다. 수9)의 짝 지워진 표본 t 검정의 경우 “...유의수준 95%...” 유의 수준의 의미를 이해하지 못했으며 회귀분석과 양의 상관계수를 혼동하고 있었다. 단위의 오류경우로 수3)의 경우에 “...430+-20% mg dry wt./1...” %단위를 비율척도 측정단위와 혼용하는 잘못된 표현을 사용하고 있는 사례이다. 수24), 수29)는 단순 실험 측정치의 비교보다는 다중비교 기법인 분산분석 기법이 필요하다.

분석결과	판정	O	X	N	소계
	논문편수	27	5	.	32
	%	84	16	.	100

정보통계그래픽스의 오용으로 수6)와 수14)의 비교 막대그래프는 세로길이/가로길이의 모양 모수가 잘못 선정되어 추세에 대한 시각적 혼동을 빚고 있다. 수6)의 경우 그래프 총수가 18개로 너무 많으며, 그래프단위는 소수점이나 설명은 %로 하고 있으며, 그래프설명 부분과 그래프표현 쪽수가 맞지 않는 등의 불일치를 보이고 있다. 그리고 Fig.7 - Fig.9의 막대그래프는 4종의 물질을 서로 비교하는 것이 불가능한 그래프이다. 또한 시계열자료에 대한 분석은 모두 경향분석의 수준에 머물고 있었다.

2) 결과해석

연구가설과 결론이 잘 일치되는 경우로 수11)“...평수기와 갈수기에 유사한 회귀계수값이 산출...”를 들 수 있다. 통계적 결론과의 일치에 있어 수1)의 경우에 “...LAS의 농도가 높았고 조사지점의 인구밀도에 비례하여 증가하는 경향이 있었다.” 인구밀도의 경우는 연구가설과 실험가설에 설정되지 않았던 부분으로 실험표본의 한 결과로 결론을 내리기 어렵다. 수21)의 경우 “...회귀분석결과 상관계수가 0.842... 자동측정의 신뢰도가 높은...”, 혹은 수28) “Fig. 6 Correlation Plot...”의 표현과 수27), 수29)에서는 상관계수와 회귀분석과의 이해부족과 용어의 오류를 보인다.

분석결과	판정	O	X	N	소계
	논문편수	28	4	.	32
	%	88	12	.	100

4.2 한국환경과학회지의 통계적 기법 활용의 타당성 평가

가. 1 단계(연구설계)

1) 연구문제와 연구가설

연구문제의 서술은 조사설계 및 실험설계의 모든 학술논문에 있어 대체로 잘 정리되어 있었다. 환 9)는 계절별 최저수질을 환경정책의 기준으로 삼는다는 연구가설을 주장하고 있으며, 환12)는 “귀무가설로 토양과 채소에 함유된 Pb의 농도사이에는 아무 상관이 없다”는 표현으로 대체로 잘 설정하고 있다.

2) 연구의 대상

분석결과	관정	O	X	N	소계
	논문편수	28	.	.	28
	%	100	.	.	100

대부분의 논문에서 모집단과 조사집단, 그 크기 규모, 특징을 대체로 정의하고 있으며, 표본채수지점의 경우 해당 구역의 지도도표를 첨부하고 있다. 한 예로 환5)에 있어 “...금호강의 영천교, 청천, 동촌...” 처럼 표본모집단을 언급하고 있다.

3) 표본설계와 실험설계

환18) “...조사지점의 선정은 환경정책 기본법 15조에 따라... 예비탐사를 거친 후...” 등 환경기본법 15조에 따라 표본지점과 실험계획이 대체로 잘 되어 있는 편으로, 환6) 물리, 화학적 실험요인 배치가 잘 설명되어 있다

분석결과	관정	O	X	N	소계
	논문편수	28	.	.	28
	%	100	.	.	100

나. 2 단계(분석 및 결과해석)

1) 통계분석기법의 선택

분석결과	관정	O	X	N	소계
	논문편수	23	5	.	28
	%	82	18	.	100

환4) 측정치의 평균값 계산으로 단순 관찰결과로만 결론을 도출하는 등 대체로 기술통계의 수준에 머물고 있다. 수량화 문제로 환2) Table.11의 경우 토양종류와 원소간의 상관계수를 계산하였으나 토양종류는 3종으로 명목척도이다. 합당한 수량화를 하지 않고서 상관관계를 계산하는 것은 무의미하다. 환18) 표2의 상관계수해석이 잘못되었다. 3개 하천 수질의 상관계수가 아니고, 그 중의 2개 하천의 상관계수가 올바른 사용법이다. 환13) “카이제곱검정은 분산의 유의성을 검정하는...” 분산비에 대한 F검정을 카이제곱 검정법으로 잘못 적용하였으며, 상대오차 상대표준편차를 혼동하여 사용하였고, 환22) Fig.4-Fig.6 회귀분석의 경우에 결정계수 값이 0.84로 계산된 것은 등분산성과 한 개의 이상치를 제거하지 않은 결과로 보인다.

그밖에 용어의 오류로 환12) 용어오류 “...t.95에서 95%신뢰수준...”은 유의수준 5%라는 표현이 정확하다.

2) 결과해석

대부분의 시계열 자료에 대해서는 단순한 경향분석 환3), 환4), 환5), 환16), 환20), 환27)에서 하고 있어 시계열 예측모형에 대한 연구가 없었다. 환10) “...지역에 비해

인근에 낚시터가 많아 ...”의 결론은 실험가설과 전혀 다른 결론을 주장하는 잘못된 결론을 내리고 있다. 환25)의 경우 식물플랑크톤 분포의 회귀식을 구하고도 상관계수만을 인용하여 결과를 유도하였으나 회귀식에 의한 추론모형으로 설명하는 것이 훨씬 타당성을 가진다. 환 9)의 연구가설은 최저치를 기준으로 한 것이었으나 실제 인용자료는 평균값 메타데이터를 인용하는 오류를 범하였다.

분석결과	판정	O	X	N	소계
	논문편수	24	4	.	28
	%	86	14	.	100

정보통계 그래픽스에 의한 결론분석으로 환21)의 식물플랑크톤 분포 산점도의 경우, 모든 경우의 구별이 없는 세로 타점 산점도를 구하였으나, 플랑크톤 종별로 구분하여, 가로로 타점하는 것이 시각적 인식의 효과가 더 높다고 할 수 있다.

5. 결론

본 연구에서 통계적 기법의 활용의 타당성 평가를 위해 1992년부터 2001년까지의 10년간 한국물환경학회지 수질보전 수록논문 491편 중 한강, 낙동강 광역수계 관련 논문 32편, 한국환경과학회지 수록논문 726편 중 한강, 낙동강 광역수계 관련 논문 28편, 한국환경분석학회 149편 중 4편을 조사한 결과, 약간의 차이는 있지만 통계적 기법 활용에 있어 상당한 문제점을 지니고 있음을 발견하였다. 제안된 점검표를 중심으로 평가 결과를 요약하고 몇 가지 개선방안을 제안하고자 한다.

1 단계 연구설계과정에서는 첫째, 연구범위를 결정하는 연구 모집단과 조사모집단에 대한 언급이 명확하게 명시되지 않고 있다. 단순히 인용논문의 연구범위, 조사모집단, 표본범위만을 이용하는 것으로 연구 모집단과 조사모집단, 혹은 비교모집단을 형성하는 표현방법은 불명확한 연구방법이다. 둘째, 채수 표본의 대표성과 실험의 범위에 대한 기술은 논문의 과학성을 보증하는 준거가 된다. 셋째, 실험계획에 있어 통제된 환경변인에 대한 언급이 거의 없다. 토지이용도, 하천 유수량의 변화, 유속의 차이, 계절변화 등과 같은 요인을 고려한 블록화, 반복에 대한 설명이 필요하다. 넷째, 측정과 분석 신뢰도가 높아야 한다. 대부분 환경기본법에 따라한다고 언급되어 있으나 기본적으로 시료측정과 분석기준의 표준이 제시되어야 한다.

2 단계 통계분석과 결과해석에서 문제되는 것은 첫째, 적절한 통계분석 기법의 선택과 적용절차 문제이다. 회귀분석을 활용한 수질보전 9편, 한국 환경과학회지 4편 경우 독립성, 정규성, 등분산성 가정을 모두 무시하였다. t 검정의 경우는 수질보전 2편, 환경과학회지 2편 모두 정규성 가정을 검토하지 않았다. t 검정을 카이제곱 검정으로 오인한 경우도 수질보전에 1편 있었다. 둘째, 통계용어의 정확한 사용이 필요하다. 회귀분석과 상관분석을 오용한 경우가 수질연구와 환경과학회지에 각 1편이 있었으며, 분산분석이후 다중비교로 t 검정을 한 논문도 수질보전에 1 편 있었다. 셋째, 사용 통계분석 패키지에 대한 언급이 부족했다. 수질보전의 경우에 3편, 환경과학회지에 2편만이 소개되었다. 넷째, 통계가설의 형식화에 모두 인색했다. 연구논문의 논리

의 비약을 막으며 분석과정과 주장을 보다 명백하게 해준다는 점에서 필요하다. 이러한 형식화의 부족은 또한 연구가설과의 연관성을 가지며 기술하는데 어려움이 뒤따른다. 다섯째, 정보 그래픽스의 경향만으로 단순히 결론을 유도해서는 곤란하다. 시계열 자료의 경우 수질보전 7편, 환경과학회지 8편 모두 단순평균 추이 그래픽만으로 결론을 유도하였다. 최소제곱법에 의한 추세직선의 산출 등 추세변동분석을 하는 것이 필요하다.

국내 한강, 낙동강 광역하천 유역 수질환경 연구논문의 통계적 기법 활용에 대한 타당성 평가를 통해 드러난 문제점과 이상복과 김말숙(1998), 최종후와 이재창(1990)의 연구결과를 참고하여 통계적 기법 활용을 위한 몇 가지 개선책을 제안한다.

첫째, 환경정보 통계교육을 위한 교육과정의 개선이 필요하다. 환경정보학부 수준에서는 환경자료를 이용한 기초통계학, 엑셀과 SAS를 활용한 정보통계학, 표본조사개론의 이수와 대학원 과정에서의 환경사례 분석을 위한 실험통계학, 통계적 분석방법론, 다변량 분석 등의 강좌를 개발하는 것이 필요하다.

둘째, 수질환경자료의 표준화와 데이터베이스 구축이다. 현재 국립환경연구소 산하 전국 각 광역 하천 수질 측정망 1844개를 운영하고 있으며 이를 위한 데이터베이스도 점차적으로 구축되어 있다. 이러한 기본측정망 자료의 활용을 통한 사전연구, 비교연구가 보다 활성화되어야 한다. 기본측정망과 더불어 개별 연구를 위한 광역하천 유역에 대한 채수, 채토 표본 지점을 확대 선정 운영하여, 지속적인 연구와 자료의 표준화가 선행되어야 하며, 이의 데이터베이스화한 자료를 활용한 양질의 연구가 계획, 추진되어야 한다. 표준표본지점 확대 선정을 위한 표본지도 제작 및 데이터베이스 설계시, 정보통계전문가, 수질 및 토양환경연구자를 활용한 공동연구가 반드시 필요하다.

셋째, 외국 이론의 무비판적 적용의 문제이다. 국내 광역하천이 가지는 환경요인은 외국의 요인과 많은 차이가 있으며, 미국 이론의 국내 하천분석 활용의 경우에 있어 가정의 검정을 할 수 있는 능력이 필요하다고 본다. 특히 QUEL2 혹은 WASP 수리모형 적용의 경우 더욱 그러하다.

마지막으로, 통계적 기법을 활용하는 연구자의 자세가 문제이다. 통계적 기법 활용과제는 적극적 실천의 문제로 과학적 이론을 확립하고자하는 연구자의 유연한 사고방식에서부터 출발한다. 학문분야별 타성에 안주하기보다는 이론과 현실의 간격을 좁히는 탐구자세를 견지하는 것이 필요하다.

참고문헌

1. 박배경, 박석순(1995), 한반도중부지역의 하천수질통계분석, 수질보전, 제 11권, 제 3호, 193-200.
2. 박석순, 박배경, 이상호(1994), 한강유역 토지이용도에 따른 지천 수질 비교, 수질보전, 제10권, 제 1호, 10-16.

3. 오경미, 조순행, 유희찬(1995), 남한강 실측치와의 통계적 비교를 근거로 한 수질 예측모형의 적용성 평가, *수질보전*, 제 11권, 제 4호, 311-321.
4. 윤기중, 안윤기, 김병수(1987), 통계의 오용과 효율적 이용에 관한 연구, *산업과 경영*, 제24권 2호, 3-37.
5. 이상복, 김말숙(1998), 통계기법 점검표 작성시 타당성평가방법, *Journal of the Korean Data & Information Science Society*, Vol. 9, No. 2, 323-336.
6. 장대홍(1995), 우리나라언론 매체에 나타나는 통계적 그래픽의 오용실태조사와, 통계적, 제도적 해결방안에 대한 연구, *응용통계연구*, 제 8권 2호, 1-26.
7. 최선희, 조현양, 최성필, 심형섭(2000), 한국환경과학회 데이터베이스 구축 및 정보화, *한국 환경과학회 추계학술회의 발표집*, 216-219.
8. 최종후, 이재창(1990), 학술논문과 통계적기법, 자유아카데미.
9. 최지용, 신은성(1997), 하천수질관리를 위한 종합수질지표의 개발과 적용, *한국 수질보전 학회지*, 제 31권, 제 4호, 415-425.
10. 한국물환경학회(1992-2001), *수질보전*.
11. 한국환경과학회(1992-2001), *한국환경과학회지*.
12. 한국환경분석학회(1998-2001), *한국환경분석학회지*.
13. International Environmetric Society(2002), *Environmetrics*.
14. White, S. J.(1979), Statistical Errors in Papers in the British Journal of Psychiatry, *British Journal of Psychiatry*, 135, 336-342.
15. www.nd.me.go.kr

[2002년 11월 접수, 2003년 1월 채택]