

상관개념의 발달과 교수학적 증재에 관한 소고

남 주 현* · 이 영 하**

이 논문의 목적은 심리학적 이론을 기반으로 처음으로 연구된 상관개념의 발달 시기와 방법을 통계학적 관점에서 고찰함으로써 교수학적 증재의 시기와 방법에 대한 시사점을 찾고자 하는 것이다. 이러한 목적 하에 심리학적 접근의 근간이 되는 Inhelder & Piaget의 1958년 연구를 심층 분석하였다. 심리학적으로 먼저 발달시기가 연구된 상관개념은 비례, 확률개념을 선결조건으로 가지며 상관계수 공식에 초점을 맞춘 형식 논리적 입장을 취하고 있었다. 그러나 통계학적인 관점에서 상관개념은 분포개념의 일부로써, 조건부 분포의 비교를 통하여 상관개념을 파악하는 수준으로부터 비례개념과 확률개념까지 적용된 조건부 확률 분포에 이르는 수준으로 다양함을 알 수 있었다. 실질적으로 문헌연구에서도 11세~12세 정도의 형식적 조작기 초반에 이르는 학습자들에게서도 조건부 분포의 비교를 통한 상관에 대한 추론이 일어나고 있었다.

이를 통해 이 논문에서는 상관개념의 교수학적 증재의 시기를 앞당기는 것에 대한 고려와 상관개념을 분포개념의 일부로써 접근하는 방법에 대한 고려의 필요성을 주장하고 있다.

1. 서 론

상관의 실생활적 의미는 매우 포괄적이어서 생각할 수 있는 모든 종류의 관계를 의미한다고 생각할 수도 있다. 옷깃만 스쳐도 인연이라거나 우주에 무엇인가 존재함을 내가 알았다는 것만으로도 그것은 나와 상관이 있다고 하는 것 등은 상관의 의미가 매우 포괄적임을 알 수 있게 한다. 그러나 통계학적으로는 어떤 현상에 대한 다른 현상의 출현 빈도를 관찰할 때 즉, 현상 A가 나타나면 현상 B가 자주 일어나거나 그렇지 않은 관계에서 상관개념이 등장한다. 이

것은 통계학에서 조건부(확률)분포와 연계되는 핵심 주제이기도 하다. 상관에 관한 추론은 일상적인 영역에서 접하는 사건을 이성적으로 이해하기 위해서나 생활 중심 소재의 과학이나 사회 영역 연구에 있어서도 중요한 역할을 한다 (McKenzie & Padilla, 1981).

상관개념의 발달과 관련된 연구는 심리학자인 Inhelder & Piaget(1955, 1958)에 의해 처음으로 시작되었는데, 상관논리를 형식적 조작기의 주요한 사고유형 중 하나로써 확률에 관한 아이디어 발달의 가장 마지막 단계로 보았다. 따라서 상관을 이해하는 것은 비례, 확률, 그리고 조합에 관한 이해를 선결조건으로 한다(Batanero et al, 1996). 상관개념의 발달시기에 관하여 통

* 이화여자대학교 대학원, joohyun@ewhain.net

** 이화여자대학교, youngha@ewha.ac.kr

계학적이 아니라 심리학적으로 먼저 연구되어 왔다면, 상관개념에 관한 심리학적 이해가 통계학적인 것과는 차이가 없는가라는 의문이 제기될 수 있다. 두 관점에서의 상관개념에 차이가 있다면 형식적 조작기가 되어야 상관에 관한 추론이 발달한다는 심리학적 결과는 상관에 관한 이해의 관점을 일부 수정한 후에도 그대로 적용 가능한가라는 의문이 제기될 수 있다.

따라서 본 연구에서는 심리학적으로 특히 Inhelder & Piaget의 이론에서 언급된 상관개념과 상관에 관한 추론의 발달과정을 통계학적인 관점으로 재해석해 봄으로써 상관개념의 교수학적 중재 시기와 방법에 관한 시사점을 새로운 관점에서 찾아보고자 한다. 이러한 목적 하에 연구문제를 정리하면 다음과 같다.

- Inhelder & Piaget의 상관에 관한 추론 연구에서 말하는 상관개념은 통계학적인 것과 차이가 있는가? 차이가 있다면 어떻게 다른가?
- 심리학적으로 밝혀진 상관에 관한 추론의 선결조건으로 비례, 확률 추론은 적절한가? 통계학적 관점에서 새로운 선결조건은 없는가?

II. 상관개념과 상관에 관한 추론

1. Inhelder & Piaget의 연구

Inhelder & Piaget는 1958년 저서 “The growth of logical thinking from childhood to adolescence”

의 15장 Random Variations and Correlations에서 상관에 관한 실험을 상세하게 다루고 있다. Inhelder & Piaget는 형식적 조작기의 논리는 논리학자가 사용하는 명제논리학과 여러 측면에서 유사하다고 주장했다(Wordsworth, 1989). 그래서 논리학자의 명제적 계산법에 기초하여 청년의 사고과정을 ‘16 이항 연산(sixteen binary operations) 체계’로 설명하려고 시도하며 이것은 특히 조합적 가능성을 다루고 있다. Inhelder & Piaget가 주장하기를 각각 참일 수도 있고 거짓일 수도 있는 두 개의 명제가 주어진다면 청소년들은 그 둘 사이에 존재하는 것으로 추론 될 수 있는 가능한 모든 논리적 관계를 체계적으로 발생시킬 수 있다(Boden, 1979). ‘16 이항 연산 체계’외에도 청년의 사고과정을 설명하기 위한 또 다른 논리적 구조 또는 모형으로 네 가지 변환 I^1 , N^2 , C^3 , R^4 을 정의하고 있다. 이것은 기초 대수학으로 규정된 일정한 변환들에 토대를 둔 것이고 두 가지 논리적 구조 또는 모형은 서로 관련된 것이다. 이것은 형식적 수준에서의 사고의 근본적인 구조들에 실질적으로 대응되기 때문에 심리학적 중요성을 가진다(Inhelder & Piaget, 1958). 이때 전도 N 은 부정을 표현하고 상보성 R 은 (반대방향으로 향하는 변환과 동치인) 대칭을 표현하고, 상관성 C 는 부정과 대칭이기 때문에, 어린 나이에 어떤 초보적인 직관이 있더라도 전도, 상보성, 상관성을 구별하고 조정할 수 있는 형식적 조작기 이전에는 실제로 이해되지 않는다고

-
- 1) 선언적 명제(disjunction) $(p \vee q)$ 는 동일변환(identical transformation) I 에 의해 자기 자신으로 변환되어서 $I(p \vee q) = p \vee q$ 가 성립한다.
 - 2) 선언적 명제 $(p \vee q)$ 는 전도(inversion) 또는 부정(negation) N 에 의해 $\bar{p} \cdot \bar{q}$ 로 변환되어서 $N(p \vee q) = \bar{p} \cdot \bar{q}$ 가 성립한다.
 - 3) 선언적 명제 $(p \vee q)$ 는 상보성(reciprocity) R 에 의해 변환되어서 $R(p \vee q) = \bar{p} \vee \bar{q} = p / q$ 가 성립한다.
 - 4) 선언적 명제 $(p \vee q)$ 는 상관성(correlativity) C 에 의해 변환되어 $C(p \vee q) = p \cdot q$ 가 성립한다.

주장한다. 이것은 비례논리, 조합논리와도 관련된 것으로 상관논리가 발달하기 위해서는 비례적 논리와 조합적 논리가 발달해야만 가능하다는 것을 전제로 하고 있는 것이다. 이런 이론적 기반을 가지고 Inhelder & Piaget는 상관논리를 출발하고 있기 때문에 형식적 조작기에 이해되는 상관을 알아보기 위한 실험으로써 2×2 분할표⁵⁾에 의한 상관개념 실험을 하고 있는 것이다. 따라서 2×2 분할표를 통해 Inhelder & Piaget가 정의하고 있는 상관개념은 명제 p , q 에 대하여 \bar{p} (또는 p^c), \bar{q} (또는 q^c)를 유도하는 전도 변환, 그 조합에 의해 네 가지 경우들을 형성하는 조합적 추론, 그리고 유도된 네 가지 경우들에서 주장을 지지하는 경우들 a^0 와 d^0 사이의 상보성(즉, $(p \cdot q) \vee (\bar{p} \cdot \bar{q})$)과 지지하지 않는 경우들 b^0 와 c^0 사이의 상보성 추론, 그리고 그런 특성을 가진 두 부류 사이에 존재하는 부정 명제적 관계성을 이해하는 과정을 통해 완성되어 간다. 이때 Inhelder & Piaget가 전제하고 있는 상관계수는

$$R = \frac{(a + d) - (b + c)}{(a + d) + (b + c)} \dots(1)$$

(단, $a = p \cdot q$, $b = p \cdot \bar{q}$, $c = \bar{p} \cdot q$, $d = \bar{p} \cdot \bar{q}$) 로써, 이것이 나타내고 있는 의미대로 상관관계를 즉각적으로 수치화할 때 가장 형식화된 수준으로 보고 있다. 따라서 Inhelder & Piaget의 관점에서는 피험자가 내재적이든 외재적이든 가설로 세우고 있는 명제에 대해 부정이나 역상사의 명제를 고려하고 각 조합에 근거하

여, a 와 d 를 합으로 결합하고 b 와 c 를 합으로 결합하여 그 차를 고려하는 수치적 반응을 하지 않는다면 상관개념이 형성되지 않았다고 보고 있는 것이다. 따라서 Inhelder & Piaget에 따르면, 확률의 마지막 단계로써의 의미를 갖는 상관에 관한 추론은 형식적 조작기 이후에서야 발달되기 시작한다.

2. 상관에 관한 추론의 선행연구들

Inhelder & Piaget 이후의 심리학적 연구들은 대부분 형식적 조작기 이후에 상관논리가 발달한다는 것에 동의를 하고 분할표를 이용하여 결합과 관련된 연구를 진행하였다. Pérez Echevarría(1990)는 심리학적 연구에서 확인된 추론들을 5개의 수준으로 요약하였다: 수준1. 표에서 단지 하나의 셀 $[a]$ 를 항상 사용한다, 수준2. $[a]$ 와 $[b]$ 또는 $[a]$ 와 $[c]$ 를 비교한다, 수준3. $[a]$ 와 $[b]$ 그리고 $[a]$ 와 $[c]$ 를 모두 비교한다, 수준4. 표 안에 있는 네 가지 셀을 가산적 비교를 수반하여 모두 사용한다, 수준5. 표 안에 있는 네 가지 셀을 승법적 비교를 수반하여 모두 사용한다. 심리학적 연구들은 분할표를 통해 상관에 관한 추론을 함에 있어서 왜 어려움을 갖는지를 연구하였다. 이러한 관점에서의 연구들은 Inhelder & Piaget가 고려한 대각선 관점의 비교- $a + d$ 와 $b + c$ 의 비교-는 한 변수에 관한 주변 도수(marginal frequency)가 똑 같은 변량들- $a + b$ 와 $c + d$ 가 똑

5) 일반적으로 통계학에서 두 개 이상의 범주형 변수에 따라 분류된 도수표를 분할표(contingency table)라고 한다. Inhelder & Piaget은 1958년 저서에서 'association table'라는 용어를 사용하고 있는데, 통계학에서의 분할표를 의미하고 있으므로 용어적 통일을 위해 본 논문에서는 분할표라는 용어를 사용한다.

6) $a = p \cdot q$
 7) $d = \bar{p} \cdot \bar{q}$
 8) $b = p \cdot \bar{q}$
 9) $c = \bar{p} \cdot q$

같은 경우-에서만 명백한 것이라는 한계점이 지적되었다(Jenkins & Ward, 1965; Shaklee & Mims, 1982; Arkes & Harkness, 1983; Allan & Jenkins, 1983). 또한 피험자의 경험과의 잘못된 결합 때문이라고 보는 입장도 있다(Chapman & Chapman, 1969; Jennings et al., 1982; Wright & Murphy, 1984; Alloy & Tabachnik, 1984). 이것은 실험과제가 가지고 있는 맥락을 주어진 자료로써가 아니라 일상생활적 경험에 근거하여 먼저 판단하게 된다는 것이다. 즉, 자료에 근거한 판단과 경험에 근거한 주관적인 판단 간에 인지적 충돌을 일으키게 된다는 것이다. 심리학적 연구들은 상관보다는 결합이라는 입장에서 사람들이 일상생활의 불확실한 상황에서 어떻게 판단을 하는가에 초점을 맞추고 있기 때문에 사람들이 어떤 현상들 간에 결합이 있다고 판단하는 근거와 원인이 무엇인가에 관심을 가지게 되는 것이다.

통계 또는 수학교육학에서의 연구들이 상관에 관한 연구를 진행할 때 대상은 모두 고등학교 이상의 성인인 것으로 보아 발달단계에 있어서 심리학적 의견에 동의하고 있는 것으로 보인다. 통계교육학자인 Garfield(2003)는 통계적 추론의 일종으로써 결합에 관한 추론을 다음과 같이 정의하고 있다: 두 변수간의 관계성을 판단하고 해석하는 방법을 아는 것, 이변량 관계성을 고찰할 때 이원분류표(2-way table) 또는 상관도(scatter plot)를 관찰하고 해석하는 방법을 아는 것, 두 변수간의 강력한 상관관계는 하나의 변수가 다른 변수의 원인이 된다는 것을 의미하지 않는 이유를 아는 것. 이렇게 정의된 이론 하에 대학생들을 대상으로 실험한 결과 분할표를 올바르게 해석하는 추론능력이 상관과 인과성을 구별하는 것보다 다소 높게 나왔다. Batanero et al.(1996)는 17-18세의 피험자들을 대상으로 분할표에서의 결합에 관한 직관

적 전략들과 선개념이 무엇인가를 연구하였다. 그 결과 분할표에서 부분적으로 옳은 것까지 포함했을 때 조건부 상대도수 분포를 비교하는 전략으로 추론을 한 비율이 평균 약 27.9%정도 이었다. 이와 비교할 때 비례를 이용한 추론은 평균 약 13.5%, 대각선에서의 도수의 합 비교를 이용한 추론은 평균 약 5.3%이었다. 즉, 피험자들이 상관에 관한 추론을 함에 있어서 조건부 확률분포의 비교를 이용한 비율이 다른 것에 비해 높았다는 것을 보여준다. 또한 피험자들은 결합의 결정론적 개념, 단일방향적 개념, 지엽적 개념과 같은 잘못된 선개념들을 가지고 있었다.

분할표를 이용한 연구 이외에도 상관도를 이용한 상관 연구들은 상관도의 측정단위(scale)때문에 피험자들이 상관관계를 과소추정 또는 과대추정한다는 연구결과를 제시하고 있다(Cleveland et al, 1982; Strahan & Hansen, 1978). 측정단위가 달라짐에 따라 산포된 자료들이 더 멀리 떨어져 보이거나 가까이 밀집되어 있거나 흩어져 있는 기울기가 가파르거나 완만하게 보이는 등의 시각적 표현변이가 이루어진다. 그러나 피험자들은 이것을 시각적 표현차제로 해석하여 두 변량들의 상관정도가 높거나 낮게 추론하게 되는 것이다. 그 외에도 일상생활에서의 상관관계를 평가하는 사람들의 능력을 연구하였을 때 피험자들이 대상 자료에 매우 친숙하고 그 자료의 등급을 매길 수 있다면 상관예측에서 상당한 정확성을 보이지만, 그렇지 않을 경우 그 반대가 된다(Kunda & Nisbett, 1986). 그러나 이 연구 결과에서도 같은 종류끼리 묶으려는 노력에 있어서 한계가 있음을 발견하였고, 실세계 상황을 이용한 권한만한 노력에도 불구하고 상관에 관한 다른 연구들처럼 여전히 현실과 동떨어져 있었기 때문에 이런 연구는 통계교사들보다 심리 측정가에게 더 유용하다(Hawkins, Jolliffe, &

Glickman, 1992).

지금까지 살펴 본 상관에 관한 대부분의 연구들은 심리학이든 통계교육학이든 그 대상은 고등학생이상의 성인이었다. 그것은 상관에 관한 추론이 확률추론, 비례추론과 관련된 것이기 때문에 형식적 조작기에 이르러야 발달 가능하고(Inhelder & Piaget, 1958), 다른 형식적 사고 능력에 비해서 형성률이 매우 낮으며 성인의 경우에도 상관관계가 없는 경우를 인식하는데 어려움이 있다(Ross & Cousins, 1993)는 심리학적 결과들에 동의하고 있는 것이라 볼 수 있다. 그렇다면 심리학적으로 먼저 연구되어 온 상관개념이나 추론의 발달에 대하여 통계학적으로도 완전한 동이가 이루어진 것인가? 본 연구는 그러한 질문으로부터 시작하기로 한다. 따라서 먼저 통계학적 입장에서의 상관개념을 고찰하고, 심리학, 특히 Inhelder & Piaget의 연구결과를 통계학적인 입장에서 재해석하기로 한다.

III. Inhelder & Piaget의 상관에 관한 추론의 통계학적 재해석

1. 분포와 상관 개념의 통계학적 의미

분포의 사전 상의 의미는 한 변수의 변량의 패턴, 한 변수의 모든 가능한 수치적 값들과 어떻게 각 값이 발생하였는지를 기록하는 것, 관찰했거나 이론적인 분류상의 발생의 빈도수를 보여주는 하나의 정렬이다(Oxford Dictionary, 2005). 다시 말해, 분포란 여러 가지 사회 현상, 자연 현상 속에서 발현되는 사물의 속성이 다양할 때 그 다양한 속성 각각의 빈도를 생각하는 개념이다. 예를 들어, 사람의 키에 대한 분포라면 속성은 ‘매우 큰’, ‘중간 정도의’, ‘매우

작은’ 등을 생각할 수 있고, 이때 ‘키가 매우 큰 사람, 매우 작은 사람’은 적고 ‘중간정도 키의 사람’은 많다는 표현처럼 어떤 속성이 더 도수가 많고 어떤 속성이 더 적은지를 아는 것이다. 속성은 키와 같이 연속적인 양에 대한 것일 수도 있고, 과일의 종류와 같은 범주적 성격을 지닌 것일 수도 있다.

7차 교육과정에서는 분포가 처음 다루어지기 시작하는 것은 1-가에서 소개되는 “분류하여 세어보기”이다. 제시된 조사 자료를 통해 한 가지 기준으로 분류하여 각 속성에 따른 도수를 세어보고, 최종적으로는 가장 좋아하는 것 또는 가장 많이 팔릴 것 등과 같이 가장 도수가 높은 속성이 무엇인지를 찾는다. 이것은 범주형 자료를 가지고 일변량 분포를 알아보고 그 분포를 최빈값으로 요약하는 것까지를 포함하고 있다. 일변량 분포에 관한 것은 표와 각종 그래프(그림 그래프, 막대 그래프, 꺾은 선 그래프, 줄기와 잎 그림, 원 그래프, 히스토그램, 도수분포다각형) 그리기를 통해 다루어진다. 기술통계학적 목표는 “분포의 특성”과 “분포의 비교”를 알아보는 것으로 자료 그 자체에 관심을 가지는 것이고, 추리통계학적 목표는 자료로부터 모집단의 특성을 추정하고 검증하는 것이다. 그러나 분포라는 범주 안에서 두 가지 입장을 모두 포함한다면, 일변량 분포에서 속성 간의 “비교”나 시간에 따른 “분포의 변화”를 살펴보는 것뿐만 아니라 변동성에도 불구하고 어떻게 미래의 상황을 예측할 수 있는가를 살펴볼 수 있을 것이다. 일변량 분포로부터 이변량 분포가 처음으로 다루어지기 시작하는 것은 상관 개념을 통해서이다. 현상 A가 나타나면 현상 B가 자주 일어나거나 그렇지 않은 관계에서 통계적 의미의 상관 개념이 등장한다. 즉, 상관은 이변량 결합(확률)분포¹⁰⁾와 관련된 것으로, 표준화된 변수들의 동일성(identity)에

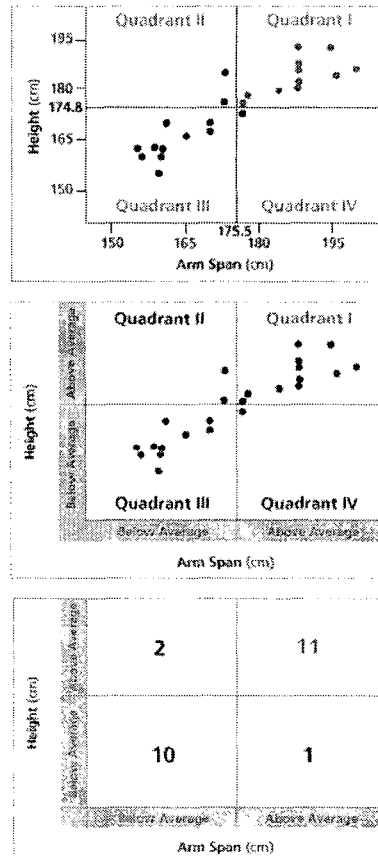
대한 근접성의 하나의 척도로 해석되어질 수도 있다(Falk & Well, 1997). 두 변수 X , Y 간의 선형적 관계성의 크기 또는 강도가 상관 계수이다. 이때 두 변수가 모두 순위 또는 숫자 변수 이라면 상관계수(correlation coefficient)는 방향성과 단위가 모두 고려되고, 명목 변수인 경우에는 상호영향을 주고받는 정도의 강도에 대한 척도로 방향성은 고려되지 않는 결합계수(association coefficient)의 측면이 된다. 변수의 성격은 차치하고 볼 때, 두 변량 X , Y 가 취하는 값이 각각 r , c 개일 때 자료구조를 <표 III-1>과 같이 구성하여 보통 나타낸다. 이때 r 개와 c 개가 표로 표현하기에 적절한 경우에만 사용된다.

<표 III-1> 결합분포와 관련된 자료구조

	x_1	x_2	...	x_c	합계
y_1					
y_2					
...					
y_r					
합계					

<표 III-1>을 대부분의 통계학 교재에서는 편의에 따라 결합 분포표라고 부르기도 하지만 특정용어로 지칭하지 않고 있다. 그러나 이것이 사건의 독립과 종속이라는 내용과 관련된 부분에서는 이런 자료구조를 $r \times c$ 분할표라고 부른다(김우철 외, 1991). 여기서는 특히 범주형 변수에 따라 분류된 것으로 각 x_i , y_j 들이 범주

가 된다. 이것을 심리학이나 교육학에서는 상관표¹¹⁾, 유관표¹²⁾ 또는 결합표¹³⁾와 같은 용어들을 사용하고 있다. 특히, 고려 중인 변수가 두 종류인 경우에는 이원분류표라고 지칭하기도 하고 (Garfield, 2003), Annenberg/CPB (2005)에서는 'contingency table'을 [그림 III-1]과 같이 상관도에서 각 사분면에 속하는 점들의 개수를 나타낸 표로 정의하기도 한다.



[그림 III-1] scatter plot과 contingency table

- 10) 엄밀하게 본다면 상관계수는 결합분포를 결합확률분포로 바꾸어 그 기대값을 가지고 계산하게 된다. 그러나 본 연구에서는 후에 논의될 피아제의 이론과 맥을 같이 하기 위해 이후에는 결합분포로만 논의하도록 한다.
- 11) 7차 교육과정에서 의한 중학교 9-나 교과서들에서는 이것을 상관표라고 부르고 있다.
- 12) Jennings et al. (1982)는 contingency table이라고 지칭하고, <Judgment Under Uncertainty: Heuristics and Biases>에서 심리학자 이영애(2001)는 이것을 유관표라고 번역하고 있다.
- 13) Inhelder & Piaget(1958)는 이것을 association table이라고 지칭하고 있다.

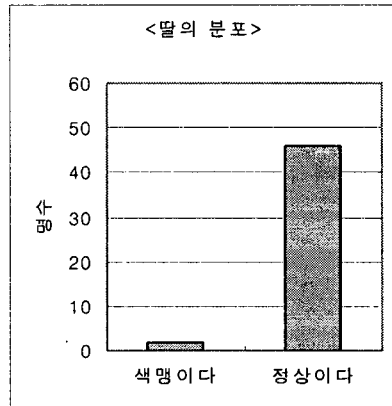
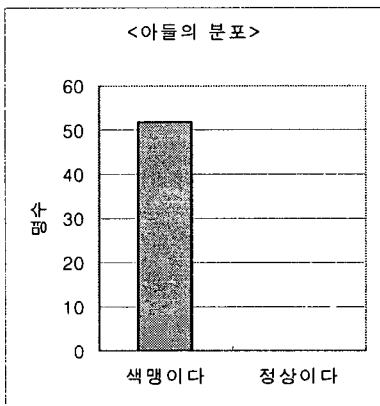
통계학에서는 수치형 자료의 경우 <표 III-1>과 같은 결합분포를 통해 조건부 기대값을 이용한다면 앞서 언급된 것과 같이 두 변량의 변화하는 방향관계와 단위까지 고려된 상관계수를 구할 수 있다. 이때 높은 상관은 두 변량의 강한 선형적 관계성을 나타낸다(Hawkins et al., 1992). 물론 통계학적으로 상관계수를 구한다는 것은 자료가 이산 또는 연속형 양적 자료라는 것을 전제로 하고 있다. 그러나 이런 결합분포를 통해 두 변량 간의 방향관계를 말할 수 없는 명목 자료와 같이 단지 두 변량 간에 상관이 있는가에 초점을 맞춘다면, 조건부 분포¹⁴⁾의 비교를 통해 직관적으로 확인이 가능하다. 예를 들어 다음과 같은 것을 살펴보자.

어머니가 색맹인 자녀 100명에 대한 조사 결과가 다음 표와 같다.

	색맹 O	색맹 X	합계
아들	52	0	52
딸	2	46	48
합계	54	46	100

이때 색맹 유전자의 발현은 성염색체와 상관이 있겠는가? (이외숙 외, 1992).

위 표에서는 색맹과 성별의 속성을 교차한 색맹 아들, 색맹 딸, 정상 아들, 정상 딸의 범주의 도수를 각각 제시하고 있다. 이때 각 도수에 근거하여 아들의 분포와 딸의 분포를 각각 그려보면 [그림 III-2]와 같다. 전문적인 통계학 지식으로 '조건부 분포'라는 것을 언급하지 않더라도 도수를 이용하여 각각의 분포를 따로 그려볼 수 있다면, 이 두 분포의 비교를 통해 색맹인 어머니의 자녀 중에서 아들은 100% 유전된 것으로 딸은 거의 유전되지 않은 것으로 추론할 수 있다. 따라서 성별의 차이가 색맹의 발현에 영향을 준다는 것, 즉 색맹 유전자의 발현과 성염색체 사이에는 상관이 있다고 할 수 있다. 이와 같은 예는 상관에 관한 추론이 형식적 확률을 강조하지 않더라도 두 조건부 분포의 비교라는 분포적 추론만으로도 가능한 경우를 보여주는 것이다. 위 예와 비교하여 다음 예를 통해 상관의 방향성이 다른 경우를 살펴보자.



[그림 III-2] 아들과 딸의 분포

14) 조건부 분포라는 것은 확률을 도입하지 않은 조건부 도수분포를 의미한다.

어떤 약이 노인들에게 소화기 문제를 유발하는지를 평가하고자 한다. 충분한 기간 동안 25명의 노인들을 대상으로 조사하였다니 다음과 같은 결과를 얻었다.

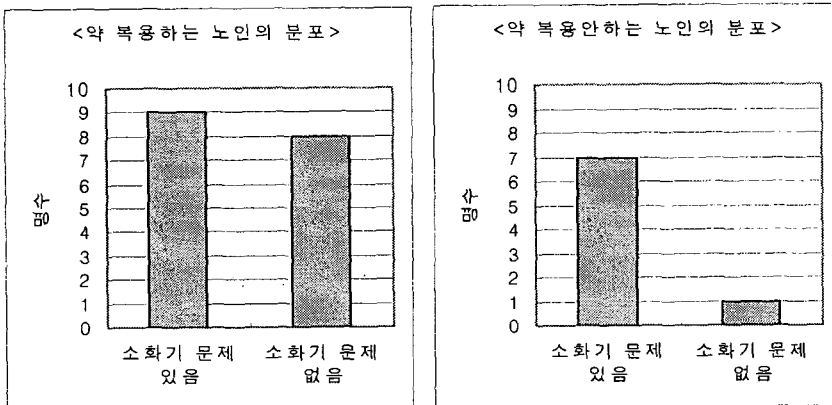
	소화기문제있음	소화기문제없음	합
약복용함	9	8	17
약복용안함	7	1	8
합	16	9	25

이 표에서 주어진 정보를 이용하여 볼 때, 노인의 소화기 문제는 약을 복용하는 것에 달린 문제라고 생각하는가?(Batanero et al., 1996)

위 예에서 약을 복용하는 노인의 조건부 분포와 약을 복용하지 않는 노인의 조건부 분포를 그려보면 [그림 III-3]과 같다. 두 조건부 분포를 비교해 보면 약을 복용하는 노인의 경우에는 소화기 문제가 있는 경우와 없는 경우 모두 높은 반면, 약을 복용하지 않는 노인의 경우에는 소화기 문제가 있는 노인의 수가 상대적으로 높음을 알 수 있다. 따라서 노인의 소화기 문제가 있고 없음은 약을 복용하지 않는 경우에 의해 영향을 받는 상관이 존재한다.

이것은 앞선 색맹과 성별의 예와 비교할 때 ‘상관이 있다’는 것은 똑같이 성립한다. 그러나 심리학에서 2x2 분할표를 통한 결합을 다루어 오면서 양적 변수가 아닌 질적 변수인 경우에도 상관의 방향성까지 언급하고 있는 것을 감안하여 구별한다면 이 두 가지 예는 상관의 방향성에서 다르다. 통계학적 의미에서 방향성은 순위 변수 또는 양적변수에서, 한 변수가 증가할 때 다른 변수가 증가 또는 감소하는가에 따라 양 또는 음의 상관관계가 있다고 한다. 그런데 심리학에서 논하는 명목변수에서의 상관의 방향성이란 두 가지 경우가 있다. 하나는 두 변수 X, Y간에 X에 따라 Y가 변화하는가 아니면 반대로 Y에 따라 X가 변화하는가의 문제로 어느 쪽이 원인이고 어느 쪽이 결과인가의 방향의 문제이다. 다른 하나는 Batanero et al.(1996)에서 언급된 것으로, 인과추론의 입장으로 상식적으로 한 변수가 증가하면 다른 변수도 증가 또는 감소해야 한다는 편견을 가지고 있는 사람이 그 생각을 자료가 “확인”시켜주는 경우를 직접적(direct) 방향으로, 그렇지 않은 경우를 역의(inverse) 방향으로 보고 있다.

이런 입장에서 보면 첫 번째 예에서 남자라는 성별이 색맹이 되는 유전자로 영향을 주는



[그림 III-3] 약 복용 노인의 분포와 약 복용하지 않는 노인의 분포

것에 비해, 두 번째 예에서 약을 복용하는 것이 노인의 소화기 질병 발생에 영향을 주는 것이 아니라 약을 복용하지 않는 것이 영향을 주기 때문에 방향성이 다르다는 것이다. 따라서 앞선 첫 번째 예와 같은 경우는 양의 상관계수를 갖는 양의 상관, 두 번째 예는 음의 상관계수를 갖는 음의 상관이라고 할 수 있다. 지금까지 두 가지 분할표의 예들을 통해 살펴본 상관은 확률개념을 굳이 적용하지 않더라도 조건부 분포의 비교를 통해 상관여부를 논할 수 있는 것들이었다. 그런데 이런 분할표를 통한 상관은 실질적으로 확률적 상관으로써 주어진 결합분포를 확률을 이용하여 결합확률분포로 바꾸어 조건부 확률(상대도수) 분포를 통해 상관을 판단하는 것이다. 또한 상관개념은 사건의 독립성과도 관련이 되며 그것은 확률개념을 필요로 하게 되어 분할표에서의 상관을 확률적 상관이라고 볼 수 있는 것이다.

상관을 사건의 독립과 종속의 내용과 관련하여 자세하게 살펴보기 위해 Batanero et al.(1996)의 연구에서 제시된 다음의 예를 살펴보자.

한 병원에서 흡연습관이 기관지 질환과 어떤 관계성을 가지는지를 판단하기 위해 250명의 사

람들을 관찰하였다. 얻어진 결과는 다음과 같다.

	기관지 질환있음	기관지 질환없음	합
흡연 합	90	60	150
흡연 안합	60	40	100
합	150	100	250

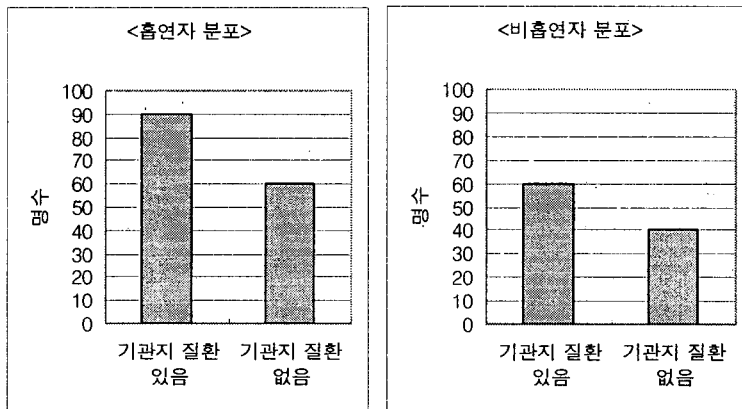
이 표에서 제시된 정보를 이용하여, 관찰된 사람들에게 있어서 기관지 질환은 흡연에 달린 문제라고 생각하는가? (Batanero et al., 1996)

2×2 분할표를 통해 앞선 예들과 같이 조건부 분포를 그려보면 [그림 III-4]와 같다. 물론 단순히 조건부 분포의 비교를 통해서만 직관적으로 상관여부를 판단하기에 혼동되는 부분들이 존재할 수 있다. [그림 III-4]의 분할표에서 흡연을 하는 사건을 A , 기관지 질환이 발생하는 사건을 B 라고 하면,

$$P(A \cap B) = \frac{90}{250} = P(A)P(B) = \frac{150}{250} \times \frac{150}{250}$$

와 같이 $P(A \cap B) = P(A)P(B)$ 가 성립한다.

두 사건 A 와 B 에 대한 수학적 관계 $P(A \cap B) = P(A)P(B)$ 는 사건 A 와 B 의 독립성을 뜻하는 것이다. 따라서 위의 예에서는 “ A



[그림 III-4] 흡연자와 비흡연자의 분포

이거나 A 가 아니거나(A^c)에 관계없이 B 와 B^c 의 상대적 빈도, 즉 비는 3:2(90:60 또는 60:40)으로 일정하다. 즉, A 의 발생여부와 무관하게 B 는 항상 $\frac{3}{5}$ 정도의 출현 확률이 있다”라고 할 수 있다. 결국 이것은 수학적 “독립성”의 의미는 언어적으로는 “상관이 없음”을 보여주는 예이다.

이와 같은 확률적 상관과는 달리 관찰하고 있는 어떤 현상을 측정된 결과들의 속성이 범주형 자료가 아니라 이산 또는 연속변수인 경우에는 다른 의미의 상관이 전개된다. 그것은 두 변량간의 선형적 관계성을 살펴보는 함수적 상관인데 여기서 관찰점들을 산점도-일반적으로 중·고등학교 교과서에서는 이것을 상관도라고 부른다-라는 그래프를 그려봄으로써 일차적으로 살펴보게 된다. 더 나아가 독립변수 (또는 설명변수)의 값이 오차없이 측정되고 실험자에 의해 조절될 수 있다는 전제 하에 임의로 고정된 독립변수의 값에 주어진 자료를 이용하여 종속변수 (또는 반응변수)의 예측을 함으로써 독립변수의 영향력을 결정하기도 한다(이외숙 외, 2002). 이것이 바로 회귀와 관련된 것으로 분포적 가정의 견지에서 보면 주어진 독립변수의 값에 관한 종속변수 상에서의 단일변량(univariate) 분포를 관찰하는 것이다(Hawkins et al., 1992). 상관과 관련된 가장 혼란 오개념은 상관과 인과성(causation)의 문제이다. 다양한 분야에서 상관에 관한 응용을 하는 주요한 목적은 우리를 둘러싼 환경을 이해할 수 있도록 하는 인과적 설명을 찾기 위해서이다. 그러나 상관은 인과적 관계성을 필수적으로 유도하지는 않는다(Batanero et al., 1996). Hawkins et al.(19

92)는 Geoffrey Rose가 사용했던 자살률 하락 예와 이것을 벗어나기 위한 의사(擬似) 상관(spurious correlation)의 예들을 언급하고 있다. 즉, 두 변인 사이의 높은 상관 결과를 인과적 종속에 있다고 생각하는 오개념을 벗어나기 위해서는 먼저 일상적 자료들이 가지고 있는 다변량적 속성¹⁵⁾을 이해해야 한다. 예를 들어 ‘언어 능력’과 ‘키’ 모두 숨어있는 제 3의 근본적 요인인 ‘나이’에 따라 증가한다는 사실의 결과로써 ‘언어 능력’과 ‘키’ 사이에 단순하게 높게 상관이 나올 수도 있다는 것이다.

지금까지 통계학적인 입장에서 상관이 크게 확률적 상관과 함수적 상관으로 나뉘어 보았다. 이 중에서 확률적 상관의 측면만 고찰한다면, 통계학에서는 실질적으로 $r \times c$ 분할표는 범주형 자료에서 사건의 독립과 종속 여부를 알아보는 조건부 (확률) 분포의 주제이다. 그러나 범주형 자료가 아닌 이산형 자료의 경우에는 분할표라는 용어보다는 결합 (확률) 분포표라는 용어로 사용되며 조건부 기대값 개념으로 상관계수를 구한다. 독립성과 상관계수의 관계는 두 사건이 독립이라면 상관계수는 0이 되지만 그 역은 성립하지 않는 것이다. 결국 통계학적 관점에서는 상관은 이변량 분포의 비교로써 조건부 (확률)분포의 비교를 다룬다는 것이 핵심이다. 따라서 상관에 관한 추론은 확률의 마지막 단계라는 심리학의 견해를 수용하더라도 앞선 예와 같이 어떤 범위 안에서는 확률개념이 없더라도 조건부 분포의 비교만으로도 상관이 있는가에 대한 고찰이 가능하다. 그것으로부터 확률개념과 비례개념이 적용되었을 때는 다른 주제까지 더 확장될 수 있는 것이다.

15) Gnanadesikan(1977)은 “자료의 많은 몸체들은 특정 배경, 환경, 또는 실험의 다양한 측면들과 결합된 관찰값들을 가진다. 그렇기 때문에 일반적인 상식으로 자료는 성격상 항상 다변량(multivariate)이다”라고 말하였다.

이제 Inhelder & Piaget의 상관에 관한 연구 결과를 통계학적인 관점으로 재해석해보기로 한다.

2. Inhelder & Piaget의 이론에 대한 통계학적 재해석

Inhelder & Piaget가 상관과 관련하여 5세부터 15세 아동에게 실시한 실험은 크게 두 가지인데 그 하나는 무작위 변동의 확률 문제이고 또 다른 것은 상관관계의 확률 문제이다. 실험과제가 과연 상관추론과 어떻게 연관되고 있는지를 살펴봄으로써 Inhelder & Piaget가 생각하고 있는 상관의 본질을 알아보려고 한다.

Inhelder & Piaget가 단계 I, II, III¹⁶⁾의 피험자에게 사용한 실험과제는 수평면에서의 공 던지기 실험과 피스톤압력과 액체 저항에 관한 실험이다. 이 실험들에서 Inhelder & Piaget가 가정한 것은 두 실험 모두 주어진 어떤 조건의 집합 하에서 엄격하게 일정하지는 않기 때문에 공의 도달지점과 액체와 피스톤에 의해 도달된 평형 단계는 확률개념을 수반한다는 것이다. 그래서 Inhelder & Piaget(1958)는 피험자가 오차변동(fluctuation)에 확률 법칙을 먼저 맞추고 오차변동에도 불구하고 운동(motion) 또는 평형(equilibrium) 실험에서의 법칙들 또는 현상의 원인들을 분리시켜야만 한다고 가정하고 있다. 이것은 두 가지 관점으로 해석이 가능하게 된다. Inhelder & Piaget가 처음 이 실험과제를 사용한 것은 운동의 보존에 관련하여 인과성의 분석을 요하는 문제 상황을 위한 것이었다. 그들은 가설-연역적 사고를 인간 지식의 정점, 즉, 과학과 수학의 지적 핵심으

로 파악한다. 그리고 이러한 형식적 조작 능력은 실험과 이론화의 논리 구조를 규정할 뿐만 아니라, 그렇게 함으로써 과학과 일상적 문제의 성숙한 이해에 중심 역할을 하는 아이의 존재하는 도식-인과성, 공간, 시간, 수, 그리고 유목화 따위와 같은-이 더욱 발전적으로 재구성되도록 영향을 미치기도 한다(Boden, 1979). 따라서 인과성의 분석을 요하는 문제에 직면했을 때 구체적 조작기의 아동과는 달리 형식적 조작기의 아동은 복잡한 인과관계를 추론해 낼 수 있다는 것이다. 이와 같이 함수적 관계성, 인과성에 기본을 두고 있는 실험은 Inhelder & Piaget의 형식적 조작기의 사고의 기본 이론과는 논리적으로 일치하더라도 상관을 인과성과 혼동시킬 여지를 줄 수 있다는 문제점이 대두된다. 그런데 통계학적으로 상관개념을 인과성과 구별되도록 명백히 하는 것은 중요한 것이다(Hawkins, 1992). 이런 중요한 점을 간과함으로써 통계학적으로 볼 때 상관의 오개념을 아동에게 줄 수 있는 것뿐 아니라, 임상실험과정에서 상관의 개연적 추론의 성격보다는 어떤 연역적 결과를 유도해나가게 되어 의미있는 아동의 추론과정이 드러나지 못했을 가능성이 있다. 그러나 달리 해석하여 Inhelder & Piaget가 현상의 원인을 격리한다는 것을 조합적 논리로만 제한하여 본다면 Wordsworth (1989)가 Inhelder & Piaget의 진자 실험설계를 <표 III-2>와 같이 해석한 것처럼, 잠재적 상관구조를 보고자 했던 단계 I, II, III에서의 실험에서도 피험자가 내적으로든 외적으로든 <표 III-2>와 같은 것을 유도할 때 Inhelder & Piaget는 상관 스키마 형성을 해 나가고 있다고 보게 된다.

16) Inhelder & Piaget는 I단계를 보존도 분포의 법칙도 없는 단계, II단계를 산만한 확률적 반응 그리고 분포의 영역 결정 단계, III단계를 가능성 오차변동(chance fluctuation)을 기반으로 하는 법칙의 분포와 결정에 관한 설명 단계로 나누고 있다.

<표 III-2> 진자 문제: 4개의 조합

	요소		결과
	길이	무게	진동
1.	길다	가볍다	느리다
2.	짧다	가볍다	빠르다
3.	길다	무겁다	느리다
4.	짧다	무겁다	빠르다

실험과제를 위와 같은 조합으로 결합시킨다면 수평면에서의 공 던지기 실험은 요소가 무게, 부피 2가지이고 무게는 3가지, 부피는 2가지로 총 조합의 수가 6가지이다. 이런 변량들의 조합을 고려한다고 했을 때 Inhelder & Piaget는 아래와 같이 이런 조합 중 피험자가 자신이 주장하고자 하는 것을 명제화하여 부정 명제를 유도하고 각 조합을 구성하게 되는 것을 상관 스키마를 형성해나가는 것으로 보고 있는 것이다.

예를 들어, 피험자는 가장 작은 공이 가장 멀리 간다는 것을 증명하길 원한다고 가정하자. p 를 고려 중인 공이 표준 공보다 더 작다는 명제라 하고 \bar{p} 는 더 크다는 명제라 하자; q 는 작은 공이 다른 것보다 더 멀리 간다는 사실을 나타내고 \bar{q} 는 지금까지 가지 않았다는 사실을 나타낸다. 변동의 관찰은 네 가지 가능한 조합의 진리를 가정하도록 피험자들을 유도한다(이는 “완전한 확인”을 위한 기호이다):

$$(p \cdot q) \vee (p \cdot \bar{q}) \vee (\bar{p} \cdot q) \vee (\bar{p} \cdot \bar{q}) = (p \cdot q)$$

이런 네 가지 가능성들은 “결합 계수”라고 불린 상관의 단순화된 표현을 생각할 때 사용된 분할표 형태에서 네 가지 셀에 대응된다는 것을 즉각적으로 알 수 있다.

	작다 (p)	크다 (\bar{p})
멀리 간다 (q)	$a = p \cdot q$	$c = \bar{p} \cdot q$
가까이 간다 (\bar{q})	$b = p \cdot \bar{q}$	$d = \bar{p} \cdot \bar{q}$

(Inhelder & Piaget, 1958: 231)

단계 III의 초기에서야 모든 가능한 조합에 관해 변인들을 분리하고 법칙을 발견하고 있으나 Inhelder & Piaget가 생각하는 상관논리에는 이르지 못하고 있다. 그러나 오히려 조합적 논리와 명제적 논리에 의한다면 상관개념에 도달되지 않았더라도 상관이 분포의 일부라는 관점에서 본다면 이러한 실험과제는 형식적 조작기 이전의 아동들에게서 분포개념의 발달을 엿볼 수 있는 결과들을 제공하고 있다. 예를 들어, 9살의 Cor는 수평면에서의 공 던지기 실험에서 연구자가 제시한 1m에는 10번을 던지더라도 도달하지 않는다는 것을 분명히 하며, 공이 갈 수 있는 범위를 10cm와 65cm라는 구간으로 예측하고 있다. 연구결과에서 제시된 응답 내용만으로는 제한이 있지만, 몇 번의 공 던지기 경험을 통해 공이 도달 가능한 거리에 대한 분포를 구체적 조작기의 아동 스스로 형성하였다는 것을 알 수 있다.

COR(9살): -“이것이 지금 어디로 갈까?[큰 알루미늄 공, 36cm].-“이전과 거의 똑같이(the same before), 여기 근처로요(around there).-“연속적으로 10번 반복한다면 여기로[1m] 갈 수 있을까?-.“아니오”-“그럼 여기는 [65cm]?”-“네”-[실험: 42,36,37,38cm] 이제 어디로 갈지 말할 수 있겠니?”-“네. 여기쯤(around there) .” [27과 47cm를 가리킨다.]-“그것은 여기로[60cm] 줄곧 올라갈 수 있었니?”-“네. 때로는(sometimes)”-“그리고 거기는 [1m] ?”-“아니오. 여기와 거기 사이만요 [10cm부터 65cm까지].” (Inhelder & Piaget, 1958, pp.227)

이렇게 분포개념이 구체적 조작기에서부터 발달되기 시작한다면 두 분포의 비교를 가능하게 하는 분포에 관한 추론은 언제부터 발달되기 시작한다고 볼 수 있는가? 예를 들어 형식적 조작기 수준에 이르지 않은 Cor에게 큰 알루미늄 공과 크기는 같으나 무게가 다른 공 던지기를 통해 똑 같은 질문을 한다면 Cor가 그

공에 대한 분포를 마찬가지로 인식하는 지를 알아볼 수 있을 것이다. 만약 인식을 한다면 두 분포의 비교를 통해 공의 무게와 거리에 대한 상관을 유도할 수 있는가에 대해 관찰가능할 것이다.

이런 실험결과들을 물론 저자들은 다른 관점에서 보았겠지만, 본 연구의 입장에서는 상관을 명제적 논리에 의해 접근하기보다는 분포의 비교로써 분포에 관한 추론의 일부로 접근해 나가는 시기와 그 방법에 대한 연구의 필요성을 시사하고 있다고 본다.

Inhelder & Piaget는 단계 III에서 무작위 변동으로부터 범칙을 분리하는 결과를 통해 이 시기의 피험자에게 잠재적 상관 스키마가 형성되어 있을 것으로 보고 단계 III-A, B의 피험자에게 '눈색깔과 머리색깔 과제'를 실험하였다. Inhelder & Piaget는 피험자가 2x2 분할표를 해석함에 있어 다음과 같은 추론들을 보이고 있음을 보고한다: 결합 a 를 독립적으로 고려하기, a 와 b 또는 a 와 c 를 비교하는 수직 또는 수평 관점, a 와 d 를 고려하는 대각선 관점, $\frac{a+d}{a+b+c+d}$ 와 $\frac{b+c}{a+b+c+d}$ 의 발견을 통해 두 차를 고려하기. 이것은 후에 Perez Echevarria(1990)에 의해 Inhelder & Piaget의 연구를 포함하여 심리학 연구에서 밝혀진 분할표와 관련된 전략들로 7가지로 요약되기도 하였다. 이러한 상관에 관한 추론 유형들이 옳고 그름의 문제는 나중에 논하더라도, Inhelder & Piaget는 명제논리학에 기본을 두고 있는 상관 논리와 (1)과 같은 상관계수 공식의 전제 때문에 상관개념의 발달을 대각선 관계성에의 고려에 초점화 하고 있는 것이다. (1)은 상관관계를 구하는 많은 공식 중의 하나이기는 하지만 일반적으로 2x2 분할표에서 표준적으로 사용되는 상관계수는

$$\begin{aligned} \varphi &= \frac{(a+b)(a+c)}{(a+b+c+d)^2} \\ &= \frac{a}{a+b+c+d} \\ &= \frac{bc-ad}{(a+b+c+d)^2} \end{aligned}$$

이다. 그렇다면 왜 합 $(a+d)$, $(b+c)$ 의 관계가 아니라 곱 ad , bc 이어야만 하는가? 합의 관계에 의한 상관계수의 입장에서 Inhelder & Piaget가 제시한 (2)~(4)와 같은 것은 $(a, b, c, d) = (4, 2, 6, 3)$ 과 같은 예를 통해서 그 문제점을 쉽게 발견할 수 있다.

$$\begin{aligned} E[(p \cdot q) \vee (\bar{p} \cdot \bar{q})] &= E[(p \cdot \bar{q} \vee \bar{p} \cdot q)] \\ &= 0 \text{의 상관관계} \dots (2) \end{aligned}$$

$$\begin{aligned} E[(p \cdot q) \vee (\bar{p} \cdot \bar{q})] &> E[(p \cdot \bar{q} \vee \bar{p} \cdot q)] \\ &= \text{양의 상관관계} \dots (3) \end{aligned}$$

$$\begin{aligned} E[(p \cdot q) \vee (\bar{p} \cdot \bar{q})] &< E[(p \cdot \bar{q} \vee \bar{p} \cdot q)] \\ &= \text{음의 상관관계} \dots (4) \end{aligned}$$

앞서 설명한 것과 같이 상관표에 의한 확률적 상관개념은 사건의 독립과 종속에 관한 내용과 관련되므로, $(4, 2, 6, 3)$ 과 같은 경우는 독립사건이기 때문에 상관관계가 없다. 그러나 위와 같은 Inhelder & Piaget의 입장으로는 $(4+3) < (2+6)$ 이 성립하므로 음의 상관관계를 가진 것이 된다. 그러므로 (1)과 같은 상관계수는 적절한 것이라고 할 수 없다(Batanero et al., 1996). 즉, 이런 대각선 관점은 Jenkins & Ward(1965)가 지적한 것처럼 조건부 빈도수가 서로 같은 경우에만 성립된다. Inhelder & Piaget는 변환의 일종으로써의 상관논리 때문에 피험자가 a 와 d 를 더할 수 있는가에 상관에 관한 추론의 발달여부 초점을 맞추지만, b 와 c 의 묶임(bond)이 a 와 d 의 묶임에 얼마나 영향

을 주는지, 어떤 경우에 b 와 c 의 묶임을 버리고 취해야만 하는지에 대한 설명은 하지 않고 있다. 이러한 Inhelder & Piaget의 상관계수에 대한 문제점과 한계를 고려하여 Jenkins & Ward(1965)는 다음과 같이 두 개의 조건부 확률의 차를 상관계수로 제안하기도 하였다:

$$\delta = \frac{a}{a+b} - \frac{c}{c+d}$$

한편 분포의 비교가 상관에 관한 추론의 핵심이라는 점을 생각한다면 곱이든 합이든 계산 과정이 중요하다기보다는 분포의 비교를 어떻게 형성해나가는지에 대한 관찰이 중요하다. 이러한 입장에서 보면 Inhelder & Piaget의 실험에서 12살 Lyn의 추론은 오히려 주목할 만하다.

그녀에게 분류할 카드를 모두 주었다. 그녀는 즉시 네 가지 속성에 따라 분류를 했다. 그리고 나서 그녀에게 다른 그룹보다 한 그룹에서 눈 색깔과 머리 색깔 사이의 관계성을 찾기가 더 쉽도록 두 개의 그룹을 형성해보도록 하였다. 그녀는 3,3,4,4와 3,6,6,4를 만들었다: “6과 4, 6과 3이 있는 반면, 3과 3, 4와 4를 가지고 있기 때문에 (3,3,4,4)에서 가능성이 더 높다.” 다른 말로 LYN이 올바르게 두 그룹을 만들었지만 그녀의 관점은 대각선 관계 $(a+d)/(b+c)$ 가 아니라 a/b 와 c/d 의 관계에 의한 추론에 의해 그것들을 조직했고 정당화했다. (Inhelder & Piaget, 1958: 236-237)

그녀가 실제로 상관이 없는 집합 3,3,4,4을 음의 상관을 갖는 집합 3,6,6,4보다 더 높은 상관을 갖는다고 한 것은 물론 잘못된 판단이다. 그러나, 주목할 만한 것은 Lyn이 비의 비교를 통한 추론을 두 집합에서 상관이 있는가를 판단하는 것에 사용하고 있다는 것이다. 학생들이 무엇을 할 수 있는가의 관점으로 볼 때, 12살의 아동에게서 비의 비교를 통한 직관적 추론이 시작되고 있다는 것은 교수학습에 중요한

단서가 될 수도 있다. Batanero와 그의 동료들이 17~18세 피험자들을 대상으로 한 실험에서 비의 비교를 통한 추론은 상관이 없는 경우, 음의 상관인 경우, 양의 상관인 경우들 각각에서 6.1%, 18.3%, 16%의 비율을 차지한다. 이와 비교할 때 12살의 아동에게서도 이런 비의 비교를 통한 직관적 추론이 이루어지고 있다는 것은 결국 Inhelder & Piaget의 입장과는 달리 형식적 조작기 이전에서도 상관에 관한 추론은 발달 가능하다는 것을 시사한다. Lyn의 비의 비교를 통한 추론은 다음 계산 과정을 통해 알 수 있듯 Jenkins & Ward (1965)가 제안한 상관계수와 본질적으로 같은 것이다:

$$\frac{a}{b} = \frac{c}{d} \leftrightarrow ad = bc \leftrightarrow ad + ac = bc + ac$$

$$\leftrightarrow a(d+c) = (b+a)c \leftrightarrow \frac{a}{a+b} = \frac{c}{c+d}$$

물론 계산을 통한 형식적 절차를 아동이 할 수 있었는지의 문제를 논하고자 하는 것은 아니다.

또한 상관표를 보고 $\frac{a}{b} = \frac{c}{d}$ 를 통해 상관관계의

여부를 확인 하였는가 또는 $\frac{a}{a+b} = \frac{c}{c+d}$ 를

통해 확인하였는가에 있어서, 어느 것이 상관개념에 있어서 더 높은 수준에서의 추론인가를 논하기는 어렵다. 다만 두 가지 추론은 수학적으로 동치명제로 비례개념에 의존하고 있기 때문에 본질적으로 같은 구조라고 볼 수 있다는 것이다. 대각선 관점의 추론이 상관여부를 판단하는데 있어서 제한점이 있기도 하지만, Batanero와 그의 동료들의 연구를 보더라도 17~18세의 피험자들도 대각선 추론을 사용하는 경우는 8% 이하에 불과하다. 비의 비교를 통한 추론이외에도 Inhelder & Piaget의 실험에서 14살 Vec의 반응은 완전하지는 않지만, 상관이 있는지를 판단함에 있어서 $\frac{c}{c+d}$ 와 $\frac{d}{c+d}$ 를 계산하여 비교

해보고, 마찬가지로 $\frac{a}{a+b}$ 와 $\frac{b}{a+b}$ 를 계산하여 비교하는 추론과정을 보이고 있다. 이것은 조건부 확률분포의 개념으로 연결되는 추론이라고 할 수 있다.

VEC(14;6)는 집합 5,1,2,4를 옳게 분류했다: “관계성이 있니?”-“글쎄요. 그런게 있긴 하네요.[그는 a와 d를 따로 빼어 놓는다]-“그래서 그것들을 빼어놓아야 하는 어떤 관계성이 있니?”-그는 b와 c를 보여준다.-“반이 있어요-아니 2/6와 4/6요.[그는 $c/(c+d)$, $d/(c+d)$ 를 계산한다.] 그리고 푸른 눈에서는 5/6와 1/6이요.”[그는 $a/(a+b)$ 와 $b/(a+b)$ 를 계산한다.]...(중간생략) (Inhelder & Piaget, 1958: 238)

Batanero와 그의 동료들은 17~18세 피험자들이 부분적으로 옳은 경우까지 포함했을 때, 상관성이 없는 경우, 음의 상관인 경우, 양의 상관인 경우에서 각각 21.6%, 37.9%, 24.2%로 조건부 상대도수 분포의 비교를 통한 직관적 추론을 하고 있음을 밝혔다. 이와 같은 조건부 확률 분포의 비교를 통한 상관여부의 판단은 응답의 종류가 2개인 2x2 분할표에 국한되는 것이 아니라, 즉, 두 변량의 응답의 종류가 각각 r개, s개일 때도 동일하게 적용될 수 있는 것이다.

예를 들어, <표 III-3>과 같은 2x s 분할표에서 $\frac{a}{a+b} = \frac{c}{c+d} = \frac{e}{e+f} = \frac{g}{g+h} = \frac{i}{i+j}$

의 여부를 살펴보고 상관을 논하는 것으로 이어질 수 있다.

2x2 분할표에서 조건부 분포의 비교를 이용한 추론을 한 피험자가 2x s 분할표, 더 나아가 r x s 분할표에서도 유사한 추론을 함으로써 상관개념을 발달시켜 갈 수 있는가에 대한 더 자세한 연구가 필요로 된다.

지금까지 Inhelder & Piaget의 I N R C 논리구조에 기초한 상관논리를 분포를 이용한 추론, 특히 조건부 분포의 비교라는 관점으로 재해석해보았다. 이러한 관점으로 보았을 때 제기된 문제점은 첫 번째로, 피아제는 상관논리를 I N R C 논리구조에 기초하여 하나의 변환으로만 보았기 때문에 상관개념에 대한 통계학적 관점과 차이를 가지게 되었다. 따라서 형식적 조작기 이전의 피험자들을 대상으로 한 실험은 변동성을 인식하고 어떤 조건 하에서 다른 하나의 변량의 분포를 보는 것이 아니라 인과성의 분석을 요하는 것으로 상관을 인식하게 할 수 있는 오류 가능성을 내포하게 되었다는 것이다.

두 번째로, I N R C 논리구조에 의해 형성된 2x2 분할표에서 대각선 관계에만 초점을 맞추고 있기 때문에, 피험자들이 보이는 비의 비교, 수직적 또는 수평적 관점의 의미에 대한 주목을 하지 못한다. 선행연구를 통해서 지적된 것처럼 Inhelder & Piaget의 대각선 상관논

<표 III-3> 2 x s 분할표의 예

	A ₁	A ₂	A ₃	A ₄	A ₅	합
B ₁	a	c	e	g	i	a+c+e+g+i
B ₂	b	d	f	h	j	b+d+f+h+j
합	a+b	c+d	e+f	g+h	i+j	a+b+...+j

리는 독립변수에 대하여 조건부 빈도수가 동일한 2×2 분할표에서만 적용이 되는 한계가 있다. 이와 달리 조건부 분포의 비교라는 관점으로 볼 때 12살 경의 피험자들이 완전하지는 않아도 비의 비교를 시작하고, 더 나아가 수직적 또는 수평적 관점으로 추론한다는 것은 중요한 점이다. 왜냐하면 이와 같은 새로운 관점의 해석은 비례적 추론과 확률적 추론이 완전히 발달된 형식적 조작기 이후에서야 상관에 관한 추론이 이루어진다는 기존의 심리학적 견해와는 다른 접근방법을 생각해보도록 요구하기 때문이다. 다시 말해, 오히려 분포에 관한 추론의 일부로써 상관에 관한 추론을 생각한다면 9살 Cor의 반응과 같이 분포개념이 형성된 시기부터 시작하여, 두 분포의 비교, 조건부 분포의 비교, 비례적 추론이 완성될 무렵의 상관에 관한 추론 등을 거쳐 조건부 확률 분포의 비교에 따른 완성된 상관 추론에 이르는 발달과정을 고찰해볼 수 있을 것이기 때문이다. 즉, 상관개념에 대한 교수학적 접근의 시기를 기존의 중학교 수준으로 미루는 것이 아니라 초등 고학년 정도의 수준으로 앞당겨볼 수 있는 가능성을 보여준다는 것이다. 물론 확률 개념이 필수적으로 필요한 수준까지를 다루는 것은 이후로 미루더라도 분포의 비교를 통한 기초적인 상관개념을 학습하는 시기를 앞당기는 것에 관한 고려가 필요하다는 것을 시사한다고 할 수 있다.

IV. 결론 및 제언

지금까지 상관에 관하여 통계학적, 심리학적 관점에서 살펴보았다. 통계학적으로 상관개념은 분포개념의 일부로써, 상관개념이 단순히 상관계수의 계산과정을 통해 이해되는 것이 아

니라 두 분포의 비교 즉, 조건부 분포의 비교에 의해 얻어지는 것임을 살펴보았다.

단순히 조건부 분포의 비교만으로도 상관을 파악할 수 있는 경우들도 존재하지만, 더 나아가 비례개념이나 확률개념을 적용하게 되면 여러 가지 경우들에서 상관여부에 대한 판단을 정확하게 해 나갈 수 있게 된다. 이러한 분포개념의 일부로써의 상관개념에 대한 관점으로, Inhelder & Piaget의 이론과 실험을 고찰한 결과, 다음과 같은 결과를 얻었다.

첫 번째, Inhelder & Piaget의 상관은 통계학적으로 상관의 일부인 확률적 상관에 국한되어 있고 $I N R C$ 논리구조를 기반으로 하고 있기 때문에 2×2 분할표를 통한 확률적 상관에 국한되어 있다. 그러나 Inhelder & Piaget가 형식적 조작기 이전의 아동들에게 행한 실험과제로 미루어볼 때 함수적 상관의 측면을 전혀 염두에 두지 않았다고 단정하기는 어렵다. 즉, 정확하게 상관개념이 무엇인가에 대한 언급이 없이 단지 명제논리학에 근거하고 있기 때문에 그들의 연구에서 불분명한 부분들이 존재한다고 볼 수 있다.

두 번째, 심리학적 관점에서는 명제논리적 측면에서 부정의 상보성이라는 변환을 상관논리로 정의함으로써, 형식적 조작기가 되어야만 발달 가능하다는 것을 논리학적 입장으로써 전제하고 있다. 그러나 상관개념의 발달은 Inhelder & Piaget의 주장처럼 비례, 확률, 조합에 관한 이해뿐만 아니라, 분포에 관한 이해도 필수적이다. 따라서 상관을 분포 비교의 관점에서 보았을 때는 11~12세의 형식적 조작기의 초반에도 충분히 발달될 수 있는 여지가 있다는 것을 앞서 그들의 실험과정에서도 찾아볼 수 있었다. 그러나 많은 선행연구들 역시 상관에 관한 관점을 확률과 비례에 관한 것으로만 집중하였기 때문에 상관에 관한 관점을 분포의 일부로 접

근하는 연구가 소홀했다. 또한 분포에 관한 연구들도 대부분 분포의 구성 또는 확인과 관련된 연구(Jones et al., 2000; Mooney, 2002) 또는 표집분포 관련 연구(delMas et al., 1999)이고, 분포 개념의 발달, 분포에 관한 추론에 대한 연구는 거의 이루어지지 않았다. 분포가 통계학에서의 중요한 개념임에도, 분포들을 비교하거나 추론하는 과제가 주어졌을 때 학생들이 종종 실패하는 것을 볼 때 학생들에게 이 개념을 발달시키는 것이 얼마나 어려운가를 알 수 있다. 그러나 통계학적 관점에서는 상관은 이 변량 분포에서 조건부(확률)분포의 비교를 다루는 주제이다. 따라서 상관 개념이 확률적 추론을 발달시키는 마지막 단계라는 심리학의 견해를 수용하더라도 앞선 예와 같이 어떤 범위 안에서는 확률 개념이 없더라도 조건부 분포의 비교만으로도 상관 여부에 대한 판단 가능성은 존재한다. 이런 경우에서부터 더 나아가 확률 개념과 비례 개념이 반드시 필요로 되는 경우까지 확장되어 간다. 따라서 상관 개념의 발달을 살펴보는 데 있어서, 확률 개념이나 비례 개념의 발달과 관련한 분포 개념의 발달을 생각해 보는 것은 중요한 의미를 갖게 된다.

다시 정리하면, Inhelder & Piaget의 상관개념은 비례, 확률, 조합에 관한 입장인 반면, 통계학적으로는 상관개념은 조건부(확률)분포로서 분포개념이 선결조건이라 점에서 가장 큰 차이가 있다. 이러한 차이점은 상관개념의 발달시기에 대한 재고의 필요성을 제기함으로써 교수학적 중재의 시기나 접근방법에 대한 논의를 가능하게 한다.

따라서 후속 연구로써 첫번째로 분포개념에 대한 명확한 정의와 속성을 밝히고 이것의 발달에 대한 연구가 필요로 된다. 상관개념이나 상관에 관한 추론을 동떨어진 주제로 보는 것이 아니라 분포개념이나 분포에 관한 추론의

일부로서 접근해 나가면서, 확률이나 비례개념과 언제 어느 시기에 만나게 되는지, 어떤 관련성을 가지는지에 대한 연구가 진행되어야 할 것이다.

두 번째로, 상관에 관한 추론의 유형에 따라 어떤 교수학적 중재로 옳은 추론으로 발달시킬 수 있는가에 대한 연구가 필요로 된다. 선행연구 결과들에서 학습자들이 보여주고 있는 상관에 관한 추론 중 부분적으로만 옳거나 옳지 않은 것들을 유형별로 정리해보면 다음과 같다. 첫번째로 특정 칸의 도수로만 추론하는 경우, 두번째로 단지 하나의 변량에 대한 조건부 분포만 비교하는 경우(예를 들어, 2×2 분할표에서 [a]와 [b]만 비교하거나 [a]와 [c]만 비교하는 경우), 세 번째로 대각선에 놓인 도수들의 합들끼리 비교하는 경우, 네번째로 단지 하나의 변량에 대한 조건부 확률분포만 비교하는 경우 등이 있다. 각 경우들이 올바른 상관 개념으로 발달되기 위해서 필요로 되는 교수학적 중재가 무엇인가를 밝히는 것이 필요하다는 것이다. Batanero et al.(1996)의 연구는 부분적으로만 옳은 대각선 관점보다 조건부 분포의 관점으로 상관을 파악하는 비율이 훨씬 더 많았음을 보여주고 있다. 이것은 Inhelder & Piaget의 실험에서 아이들이 직관적으로 옳게 추론해 나가고 있었음에도 불구하고 잘못된 목표 설정으로 전반적으로는 상관개념 자체의 교수학습 시기가 더 늦어진 것은 아닌가라는 의문을 제기하게 하는 것이다. 그렇기 때문에 적절한 교수학적 중재만 이루어졌다면 그 시기를 더 앞당길 수 있지 않음을 입증하고자 하는 것이다.

세 번째로, 지금과 같이 상관개념에 대한 교수학습 시기를 비례개념이나 확률개념이 발달하는 15~16세 경의 형식적 조작기 이후에나 가능하다고 생각하고 중학교 3학년 과정으로 미

루거나 대학교육으로 넘겨 버리게 된다면, Jennings et al.(1982)의 연구결과처럼 서로 상관성이 있는가를 주어진 자료를 통해 가능한 한 객관적인 관점으로 접근하기보다는 자신의 일상생활적 경험에 근거한 주관적 판단에 의존할 우려가 있다. 확률개념이 필수적으로 필요한 수준까지를 다루는 것은 확률개념이 발달된 이후로 미루더라도, 분포의 비교를 통한 기초적인 상관개념의 도입 시기는 중학교 3학년보다 더 앞당겨서 순차적으로 학습하는 것에 관한 고려가 필요하다. 따라서 분포개념에 관한 연구와 그 맥락 속에서의 상관개념에 대한 연구를 통해 총체적인 관점으로써의 교수학습 설계를 재고할 필요가 있다. 즉, 상관개념의 발달을 위해 그 이전의 발달단계의 학습자들에게 어떤 교수학습과정이 미리 준비되어야 하는지를 밝혀야 할 것이다.

참고문헌

- 강육기 외(2002). *중학교 수학 9-나*. ㈜두산.
- 김우철 외(1991). *현대통계학*. 영지문화사.
- 이외숙 외(1992). *통계학*. 경문사.
- Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation, 14*, 381-405.
- Alloy L. B., & Tabachnik N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review, 91*, 149.
- Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112*(1), 117-135.
- Batanero, C., Estepa, A., Godino, J., & Green D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*(2), 151-169.
- Boden, M. (1979). *피아제*, (서창렬, 역). 시공사.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271-280.
- Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science, 216*, 1138-41.
- delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3).
- Falk, R., & Well, A. D. (1997). The many faces of the correlation coefficient. *Journal of Statistics Education, 5*(3), 1-18.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*(1), 22-38.
- Hawkins, A. Jolliffe, F., & Glickman, L. (1992). *Teaching statistical concepts*. Harlow: Longman Group UK Limited.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge & Kegan Paul Ltd.
- Jenkins, H. M., & Ward, W. C. (1965).

- Judgement of contingency between responses and outcomes. *Psychological Monographs*, 79, (1, Whole No. 594).
- Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments, In Kahnemann, Slovic, Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, 이영애(역), 불확실한 상황에서의 판단: 추단법과 편향, 아카넷.
- Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning* 2, 269-307.
- Kunda, Z., & Nisbett, R. E. (1986). Prediction and the partial understanding of the law of large numbers. *Journal of Experimental Social Psychology*, 22, 339-354
- McKenzie, D. L. & Padilla, M. J. (1981). Patterns of reasoning: Correlational thinking. *Paper resented at the 54th annual meeting of the National Association for Research in Science Teaching, Ellenville, NY.*
- Mooney, E. S. (2002). A framework for characterizing Middle School Students' statistical thinking. *Mathematical Thinking and Learning* 4, 23-63.
- Oxford Dictionary. (2005).
- Pérez Echevarría, M. P. (1990). *Psicología del razonamiento probabilístico*. Madrid: Ediciones de la Universidad Autonoma de Madrid.
- Ross, J. A., & Cousins, J. B. (1993). Patterns of student growth in reasoning about correlational problems. *Journal of educational psychology*, 85(1), 49-65.
- Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of experimental Psychology: Learning, Memory, and Cognition*, 8, 208-224.
- Strahan, R. F., and Hansen, C. J. (1978). Underestimating Correlation From Scatterplots. *Applied Psychological Measurement*, 2, 543-594.
- Wordsworth, B. J. (1989). Piaget's theory of cognitive and affective development, Longman Pub Group.
- Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General*, 113, 301-322.
- Annenberg/CPB. (2005). Learning Math: Data Analysis, Statistics, and Probability <http://www.learner.org/resources/series158.html>

The Development and Didactic Mediation of the Correlation Concept

Nam, Joo Hyun (Graduate School of Ewha Womans University)

Lee, Young Ha (Ewha Womans University)

The purpose of this study is to find out the implications on when and how the correlation concept can be taught. we investigate the development time and method of the concept in a statistical perspective those initially have discussed in psychology by Piaget.

We first reviewed the 1958 research by Inhelder and Piaget. It was the first one which researched the development of the correlation and has become the foundation of psychological perspective.

According to them, the correlation concept needs proportional and probability concept ahead of its development and argued on the coefficient of correlation based on formal and logical position.

However, from a statistical perspective,

the correlation concept is a part of the distribution concept. So, the level of the correlation concept grows from the comparison of conditional distributions to the conditional probability distribution where the proportional concept and probability concept are applied.

As reviewed through the literature, we found that 11-12 years old students in early formal operation stage reasoned about correlation through the comparison of conditional distributions.

In our study, we argue that we need to consider the possibility of beginning didactic mediation for correlation concept earlier and the method approaching it in a distribution perspective.

* **Key words** : correlation(상관), distribution(분포), conditional distribution(조건부 분포), conditional probability distribution(조건부 확률 분포), Inhelder and Piaget

논문접수 : 2005. 5. 22

심사완료 : 2005. 8. 1