經營學硏究 第23卷 第4號(1994年 11月)

상호작용효과를 포함한 다중회귀분석에서 주효과의 검증에 대한 연구*

Testing Main Effects in Interactive Multiple Regression

이 유 재**

최 초 논문접수일 : 94. 1
수정본 접 수 일 : 94. 6
논 문 게재확정일 : 94. 7

초 록

본 연구는 회귀분석에서 구간척도로 측정된 변수들의 주효과와 상호작용 효과를 검증할때 제기되는 문제점들을 살펴보고 그 해결책을 제안하고자 한다. 회귀모델에 상호작용 항목을 삽입하면, 주효과 평가에 있어 다음과 같은 문제들이 발생한다. 첫째, 척도의 선형변환에 의해 결론이 상당히 다르게 나타날 수 있다. 둘째, 심각한 다중공선성 (multicollinearity)이 발생한다. 세째, 주효과에 대한 해석이 애매모호해진다. 이 문제들에 대한 해결수단으로 소집단 분석 (sub-group analysis), 독립화 변환 (orthogonal centering), 평균변환 (mean centering)등을 제시하고, 이들에 대한 비교를 하였다. 그 결과, 평균변환이 이 문제들을 해결하는 데 있어서 가장 유용한 도구임을 발견하였다.

* 이 논문의 초고에 대해 유익한 제언을 해주신 두 사독위원께 감사드린다.
** 서울대학교 경영대학 교수

Ⅰ. 서 론

마케팅을 비롯한 여러 행동과학에서는 두가지 혹은 그 이상의 변수들의 주효과(main effect)이외에 그들사이의 상호작용 효과(interaction effect)에 대해서도 검토를 하는 것이 보통이다(Laroche and Howard,1980). 만약, 종속변수(Y)와 독립변수(X_1, X_2)가 계량적 척도에 의해 측정이 되었다면 주효과와 상호작용 효과는 다음과 같은 회귀모델을 통해 검토될 수 있다(e.g., Cohen and Cohen, 1983;Lilien and Kotler,1983).

$$Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_1 X_2 + e \qquad (1)$$

이때, B_0는 Y절편을 나타내고 B_i(B_1, B_2, B_3)는 회귀계수들을 나타낸다. e는 잔차(residual)를 의미한다. 일반적으로 X_1과 X_2가 상호작용을 한다는 것은 X_1과 Y의 관계가 X_2의 값에 따라 변화한다는 것을 의미한다. 이와 같은 상호작용 효과는 X_1과 X_2의 곱인 X_1X_2 변수를 회귀모델에 도입하여 측정할 수 있다. 이 경우 X_2가 X_1의 Y에 대한 효과를 조정한다고 하며 X_2를 X_1과 Y의 관계에 있어서 조정변수(moderator variable)라고 부른다. 마찬가지로 X_2와 Y의 관계에 있어서는 X_1이 조정변수가 되는 것이다. 위 등식은 수집된 자료에 어떠한 변환도 하지 않은 상태에서 제시되었음므로 기초모델(raw score model)이라 부르기로 하자. 이 기초모델에서 각 독립변수들의 효과를 검토하기 위해서는 각 회귀계수들의 유의도를 검증하게 된다. 즉, B_1의 유의도는 종속변수 Y에 대한 독립변수 X_1의 주효과를 검토하기 위한 것이며 같은 논리로 B_2의 유의도는 X_2의 주효과를 검토하기 위한 것이다. 반면에, B_3는 X_1과 X_2의 상호작용 효과를 측정하기 위해 검토하게 된다. 만약 B_3이 유의적이면 상호작용 효과가 존재한다고 결론을 내릴 수 있다. 그리고 B_3의 부호가 正(+)이면 X_1의 효과는 X_2가 증가함에 따라 커지므로 상호작용 효과가 시너지효과를 보인다고 볼 수 있다. 반면에 B_3의 부호가 負(−)이면 X_1의 효과는 X_2가 증가함에 따라 감소함으로 X_2가 X_1의 효과를 억제한다고 볼 수 있다.

이와 같은 접근은 독립변수가 비율척도(ratio scale)가 아닌 구간척도(interval scale)로 측정되었을 경우, 몇가지 문제점을 가지게 된다(Bagozzi,1984; Yi, 1989). 첫째, 구간척도에 허용가능한 변환을 통해 척도의 원점이 이동하면, 같은 자료를 대상으로 분석하였지만 독립

변수(X_i)의 주효과에 대해 결론이 달라질 수 있다. 예를 들어, 척도상에 원점을 어떻게 잡았는가에 따라 강하게 제시된 주효과가 전혀 효과가 없는 것으로 분석될 수 있다. 둘째, 회귀계수의 추정에 있어 정확도가 낮아져 중요한 예측변수의 유의한 효과를 파악하지 못하게 될 수 있다. 세째, 주효과의 해석이 곤란해진다. 많은 측정이 구간척도를 통해 이루어진다는 것을 고려한다면 이와 같은 문제점들은 매우 중요한 의미를 가지며 이 문제점들을 올바로 이해하고 해결하는 것이 필요하다.

본 연구는 다중회귀분석에서 구간척도로 측정된 변수가 갖는 주효과와 상호작용 효과를 평가할때 발생하는 이와 같은 문제점들을 검토하려고 한다. 이전에 많은 연구들이 상호작용 효과를 평가하는 문제를 검토하였다(e.g., Allison,1977; Bagozzi,1984; Holbrook,1977). 그러나, 다중회귀분석에서의 주효과를 평가하는 데 있어 제기되는 여러 이슈들을 검토한 연구들은 그리 많지 않은 실정이다(Yi, 1989). Yi(1989)의 연구에서는 상호작용이 포함된 다중회귀분석에서 발생하는 문제점들을 간단히 지적한 바 있다. 본 연구는 Yi(1989)의 연구를 다음과 같은 측면에서 확장하고 있다. 첫째, 발생된 문제점들에 대한 해결책의 여러가지 대안을 제시하고 체계적으로 검토한다. 독립화 변환이나 소집단 분석 등을 평균변환과 함께 검토하여 상대적인 강약점을 파악하려고 하였다. 둘째, 문제점의 토론이나 해결수단의 제시에 있어서 구체적으로 수학적 이론적 근거를 제공하였다. 세째, 종합적인 실증분석의 예를 제공함으로써 각 문제점들과 해결수단들을 보다 포괄적으로 예시하였다. 척도종속성, 다중공선성, 해석의 애매성 등 문제점들을 여러가지 해결수단마다 비교검토하여 문제점에 대한 보다 체계적인 이해를 제공하고 어느 해결수단을 채택해야 하는가에 대한 결정에 도움이 되도록 하였다.

본 연구는 이와 같은 이슈들을 체계적으로 분석하고자 한다(e.g., Althauser, 1971 ; Lane, 1981). 이 논문은 다음과 같은 순서로 구성되어 있다. 우선은 주효과와 상호작용 효과들을 검토하는 데 있어 제시되는 문제점들을 확인하고 이에 대한 몇가지 해결수단들을 살펴보기로 한다. 본 연구에서 제안하고 있는 해결수단은 평균변환, 독립화 변환, 소집단 분석등이며 이들의 강약점을 비교하면서 어느 변환이 다른 수단에 비해 우월한가를 살피고자 한다. 또한, 이와 같은 문제점들을 관련된 실제 자료들을 분석하며 예시하고 각 해결수단들을 비교/검토하기로 한다.

Ⅱ. 주효과와 상호작용 효과의 검증에 있어 제시되는 문제점

여기에서는 회귀모델에서 구간척도로 측정된 변수들이 갖는 주효과와 상호작용 효과를 검토하는 경우, 발생하는 이슈들을 검토하기로 하자. 특히, 척도 종속성(scale dependence), 다중공선성(multicollinearity)이란 문제들과 효과들에 대한 의미있는 해석이란 세가지 이슈들을 집중적으로 고려해 보기로 하자.

1. 척도 종속성

변수가 비율척도가 아닌 구간척도로 측정되는 경우, 척도의 원점은 임의로 정해진다. 임의적인 원점은 존재하지만, 해당 속성이 전혀 없는 상태인 절대적인 원점은 존재하지 않는다(채서일,1993) 즉, 일정한 상수를 가감해도 서열순위와 등간결과와 같은 구간척도의 속성을 유지할 수 있다는 의미이다. 예를 들어, 7점척도는 1에서 7까지로 구성될 수 있지만 0에서 6으로도 구성할 수 있다. 실제로 마케팅 연구에서 기대가치 모델(expectancy-value model)을 이용할때 7점척도로 측정한 항목을 1에서 7까지의 점수를 주거나 또는 −3에서 +3까지의 점수를 주고있다(Bagozzi,1984; Holbrook,1977, Ryan and Bonfield,1975). 그런 식의 변환이 이론적으로 가능하기 때문에 변수의 효과에 대한 결론에 아무런 영향을 주지 않아야 할 것이다. 실제로, 상호작용 효과에 대한 결론은 위와 같은 변환하에서는 변동되지 않는다. 그러나, 주효과에 대한 결론은 사용된 원점이 무엇이냐에 따라 상당히 상이하게 나타날 수 있다.

이와 같은 특성들은 다음과 같이 보일 수 있다. 우선, X_1과 X_2의 원점을 바꿔 새로운 변수를 구해 이를 X^*_1과 X^*_2라 정의하자. 즉,

$$X^*_1 = X_1 - c,\ X^*_2 = X_2 - d \qquad \text{단, c와 d는 임의의 상수임.}$$

이것은 X_1의 원점을 c만큼 이동한 것이며, X_2의 원점을 d만큼 이동시킨 것이다. 위의 식을 기초로 X_1과 X_2에 대해 풀고 이를 식 (1)에 결과를 대입하면 다음과 같은 결과를 얻게 된다.

$$Y = (B_0 + cB_1 + dB_2 + cdB_3) + (B_1 + dB_3)X^*_1 + (B_2 + cB_3)X^*_2 + B_3X^*_1X^*_2 + e \quad (2)$$

그러면 X_1과 X_2의 원점을 이동시켜 얻게 된 결과는 무엇인가? 새롭게 정의된 X^*_1의 회귀계수들이 원점의 이동에 의해 영향을 받았다는 것을 확인할 수 있다. 예를 들어, X_1의 측정척도를 선형변환시키면 X^*_2의 회귀계수가 변화된다. 따라서, X_1의 원점을 변화시켜 X^*_2의 계수를 0으로 만들 수도 있다는 것을 알 수 있다. 만약, 임의의 상수 c를 $-B_2/B_3$로 가정한다면 X^*_2의 주효과는 0이 된다. 즉, $X^*_1 = X_1 + B_2/B_3$라는 변환에 의해 X_1의 원점을 $-B_2/B_3$ 만큼 이동시키는 경우, X_2의 주효과는 사라지게 된다. 같은 논리가 X_2에도 적용이 된다. X_2의 원점을 d만큼 이동시키면 X_1의 회귀계수가 B_1에서 $(B_1 + dB_3)$로 바뀌게 된다. 즉, X_2의 원점을 변화시키면 X^*_1의 회귀계수가 변화된다는 것이다. 결국, 상호작용 효과를 고려할 경우, 한 독립변수(X_2)의 척도를 선형변환시켜 다른 독립변수(X_1)의 주효과에 대한 결론을 바꿀 수 있다는 것이다. 또한, 이 경우에 표준계수와 그들의 t값들도 함께 변화하게 된다 (Cohen, 1978). 정리하면, 한 독립변수의 원점을 변화시키면 다른 독립변수의 회귀계수가 영향을 받게 되고 그에 따라 해당 변수의 주효과에 대한 결론이 달라지게 된다는 것이다.

반면에, 상호작용을 나타내는 $X_1 X_2$항의 계수는 위와 같은 변환에 의해 영향을 받지 않는다. 식(1)과 식(2)를 비교해 보면, 상호작용항의 계수는 독립변수의 선형변환에 의해 바뀌지 않는다는 것을 확인할 수 있다. 물론, $X_1 X_2$항의 다른 변수와의 0차(zero-order) 상관관계는 척도에 따라 바뀌지만 편회귀계수(partial regression coefficient)와 그에 따른 표준오차 (standard error)는 척도의 변화에 따라 변동되지는 않는다(Allison, 1977). 결과적으로, 상호작용 효과에 대한 유의도검증은 척도의 원점을 변화시킨다고 해서 달라지지 않는다.

그러나, 회귀등식을 $X_1 X_2$항과 같은 곱셈항만을 사용하고 X_1과 X_2와 같은 덧셈항을 제거하여 구성하는 경우, 상호작용 효과에 대한 계수와 유의도검증은 선형변환에 따라 변동된다. X_1X_2항만을 가지고 있는 회귀식(즉, $Y=a+bX_1X_2$)에서는 $X_1 X_2$항의 계수가 상호작용 효과 뿐만 아니라 X_1과 X_2에 의해 선형적으로 설명되는 분산 즉, 주효과를 함께 보유하고 있다. 상호작용에 대한 적절한 검증작업에는 X_1과 X_2의 주효과들에 대한 통제가 필요하다 (Cohen, 1978). 즉, 주효과의 영향을 통제하고 남는 부분이 상호작용 효과이기 때문에 상호작용 효과를 검증하기 위해서는 식 (1)과 같이 주효과를 나타내는 항들을 회귀모델에 삽입해야 한다.

식 (1)과 식 (2)는 수학적으로 동일하기 때문에, 두 식의 예측적 정확도(predictive accuracy)는 같다(Cohen, 1978). 또, R^2에 대한 통계적 유의도를 나타내는 F검증은 R^2, 독립

변수의 수, 표본크기에 따라 결정되므로 F값은 두 식에서 같게 나온다. 결과적으로, 척도의 원점을 변화시키는 것에 의해 상호작용 효과와 모델의 예측적 정확도가 아무런 영향을 받지 않음을 확인하였다. 그러므로, 본 연구에서는 척도변환에 의해 영향을 받는 주효과의 추정과 해석에 초점을 맞추기로 하자.

이런 척도 종속성의 문제가 실증연구 등에서 종종 무시되어 온 것이 사실이다(Laroche and Howard,1980; Hornik,1984). 주효과의 척도변환에 대한 민감성은 주효과와 상호작용 효과에 대한 회귀분석에 잠재적으로 어려움을 초래한다. 한 독립변수가 비율척도가 아닌 구간척도를 통해 측정되는 경우, 다른 변수의 회귀계수의 이론적 의미는 사라지게 된다. 왜냐하면, 구간척도로 측정되는 독립변수에 상수를 가감함으로써 다른 독립변수의 주효과를 얼마든지 조작할 수 있기 때문이다. 원점변화에 따라 주효과가 正(+), 負(−), 또는 0이 될 수 있다. 그러므로, 변수의 선형변환에 의해 변동되지 않는 결론을 끌어낼 분석도구가 요구된다.

2. 다중공선성 (multicollinearity)

보통, 상호작용 효과를 나타내는 곱셈항($X_1\ X_2$)은 그 구성요소들 (X_1과 X_2)과 상관관계를 가지게 된다. 이 상관관계의 크기는 X_1과 X_2의 평균, 표준편차(standard deviation), 상관관계등에 의해 영향을 받는데, 현실에서는 종종 큰 상관관계가 나타나고 있다(Althauser, 1971 ; Tate, 1984). 특히, X_1과 X_2의 평균들이 0과 다르면 다를수록 회귀식은 더 높은 다중공선성을 보인다.

이와같은 사실은 다음의 식에 의해 보일 수 있다. X_1X_2항과 X_1항, X_2항 간의 공분산은 다음과 같다.

$$COV(X_1X_2,\ X_1) = E(x_1^2x_2) + V(X_1)E(X_2) + COV(X_2,\ X_1)E(X_1)$$

$$COV(X_1\ X_2,\ X_2) = E(x_1x_2^2) + V(X_2)E(X_1) + COV(X_1,\ X_2)E(X_2)$$

여기서 x_1, x_2는 각 변수의 평균변환 값으로 $x_1 = X_1 - \mathrm{E}(X_1)$, $x_2 = X_2 - \mathrm{E}(X_2)$이다. 여기서 $V(X_1)$은 X_1의 분산이다. 부록 1에 이 식들의 도출과정이 나와 있다. X_1과 X_2가 정규분포를 따른다고 가정하면 모든 홀수차의 모멘트는 0이다(즉 $E(x_1) = E(x_1^2x_2) = E(x_1x_2^2) = 0$). 따라서 정규분포를 가정할 경우, 다음과 같이 표현될 수 있다.

$$COV(X_1X_2,\ X_1) = V(X_1)E(X_2) + COV(X_2,\ X_1)E(X_1)$$

$$COV(X_1X_2,\ X_2) = V(X_2)E(X_1) + COV(X_1,\ X_2)E(X_2)$$

예를들어 X_1과 X_2의 공분산이 0보다 큰 경우, X_1과 X_2의 평균이 0과 차이가 나면 날수록 공분산과 상관관계가 높아지게 된다. 이와 같은 X_1과 X_2의 분포에 대해 가정을 하는 경우나 아무런 가정을 하지 않는 일반적 경우에 모두 적용된다.

다중공선성은 회귀계수의 추정과 해석에 여러 문제들을 야기시키는 것으로 알려져 있다(e. g., Belsley et al., 1980). 즉, 추정의 표준오차(standard errors)가 매우 높아지고 계수들의 t통계량은 매우 낮게 된다. 이것은 다음의 식에 의해 보일 수 있다.

$$Var(B_i) = (1 - R^2)/(1 - R_i)(N - P - 1)$$

여기서 R = Y와 X_i들 간의 다중상관계수

R_i = X_i와 다른 $X_j (j \neq i)$들 간의 다중상관계수

N = 표본크기

P = 독립 변수의 수이다.

R, N, P가 주어졌을때 추정치 B_i의 분산은 독립변수들(즉, X_1, X_2, $X_1\ X_2$)간의 다중공선성을 나타내는 R_i가 증가함에 따라 커진다.

결과적으로, 종속변수와 독립변수들 전체간에 명백한 통계적 관계가 존재하는 경우에도 각 독립변수에 대한 개별적 회귀계수는 통계적으로 유의하지 못한 결과를 보이게 되는 경우도 있다. 각 추정치의 표준오차가 비정상적으로 커지게 되므로 생기는 현상이다. 또 다른 문제는 추정치와 각 추정치의 표준오차들이 매우 불안정하다라는 것이다. 자료의 측정오차나 반올림 등으로 인한 작은 변화에 의해서도 추정치와 그 표준오차에는 커다란 변화가 야기될 수 있다. 또한, 각 회귀계수의 해석도 곤란해진다. 정리하면, 상호작용 효과를 포함하는 회귀분석에는 다중공선성이 존재하며 그 결과, 부정확하고 불안정한 결론을 유도할 수도 있다는 것이다.

3. 주효과의 해석

상호작용효과가 있는 다중회귀분석에서 한 독립변수의 주효과를 평가하기 위해서는 그 변

수의 회귀계수를 검토하게 된다. 식 (1)을 예로 볼 때, X_1의 주효과는 B_1에 의해 파악된다. 그러나, 그 회귀계수의 해석은 어떻게 해야 하는가? 이를 위해서는 해당 변수의 한계효과(marginal effect)를 검토해야 된다. 예를 들어, X_1의 한계효과를 확인하기 위해 회귀식의 편미분을 이용할 수 있다 ($\partial Y / \partial X_1 = B_1 + B_3X_2$). 이 경우, B_1은 다른 독립변수(X_2)가 0인 경우 종속변수 Y에 대한 X_1의 효과가 된다. 즉, 주효과는 해당 변수와 상호작용하는 다른 변수가 0의 값을 가지는 경우에 그 변수의 효과를 의미하게 된다.

식 (1)은 다음과 같이 Y와 X_1과의 식으로 바꾸어 볼 수 있다.

$$Y = (B_0 + B_2X_2) + (B_1 + B_3X_2)\ X_1 + e$$

즉, 식 (1)은 Y절편이 $(B_0 + B_2X_2)$이고 기울기가 $(B_1 + B_3X_2)$인 회귀식들의 집합으로 표현된 것이다. 여기서 Y와 X_1의 관계를 보면 Y절편이나 기울기가 X_2의 값에 따라 달라지는 것을 알 수 있다.

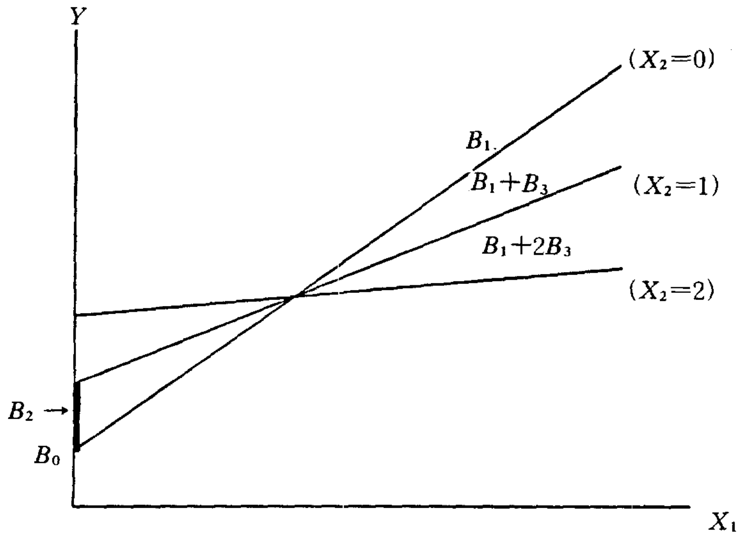
구체적인 예를 몇 개 보면 다음과 같다.

$X_2 = 0$일때	Y절편 $= B_0$,	기울기 $= B_1$
$X_2 = 1$일때	Y절편 $= B_0 + B_2$,	기울기 $= B_1 + B_3$
$X_2 = 2$일때	Y절편 $= B_0 + 2B_2$,	기울기 $= B_1 + 2B_3$
$X_2 = 3$일때	Y절편 $= B_0 + 3B_2$,	기울기 $= B_1 + 3B_3$

이것을 그림으로 나타내보면 [그림 1]과 같이 X_1과 Y의 관계가 X_2의 값에 따라 변화하는 것을 보이는 것이다. 이 그림은 $B_1 > 0$, $B_2 > 0$, $B_2 < 0$인 경우를 가정한 후에 그린 것이다. 또 식 (1)에서 X_1의 회귀계수인 B_1은 다른 변수인 X_2가 0일때 X_1의 효과를 의미하는 것을 확인할 수 있다.

이와 같은 해석방법에 깔려있는 한가지 가정은 주효과란 다른 변수들의 효과가 존재하지 않는 경우에 그 변수의 효과라는 것이다. 그러나, 그런 상황은 매우 비 현실적이다. 왜냐하면, 다른 변수들의 영향이 전혀 없는 경우가 없기 때문이다. 이 해석방법의 또 한가지 가정은 다른 모든 변수가 0의 값을 가진다는 것이다. 그러나, 모든 척도에서 각 변수가 0의 값을 가질 수는 없다. 예를 들어, 어떤 척도는 1에서 7까지의 점수로 구성될 수 있다. 즉, 0의 값이 변수가 취할 수 있는 영역밖에 위치할 수 있다는 것이다. 그런 경우에, 한 변수의 주효과는 다른 변수가 존재하지 않는 지점에서 그 변수의 효과를 의미한다. 비록 원점이 척도상 존재하는 경

[그림 1] X_2의 값에 따른 X_1과 Y의 관계



우에도 원점이 실제적으로는 무의미하고 비현실적일 수 있다. 예를 들어, 운동선수의 체력을 신장과 몸무게를 가지고 예측한다고 보자. 이런 경우에 신장변수에 해당하는 회귀계수는 몸무게가 0Kg인 운동선수에 있어서 신장이 체력에 미치는 영향을 나타낸다. 그러나 몸무게가 0Kg인 운동선수는 존재하지 않으므로 이와 같은 계수가 의미가 없는 것은 자명하다고 볼 수 있다. 이런 비현실적인 가정들이 내포되어 있기에 이와 같은 해석방법의 적용가능성은 그만큼 줄어든다고 하겠다. 상호작용 효과를 고려하는 모델은 한 변수의 효과가 다른 변수의 효과와 함께 변동한다는 것을 가정하기에, 당연히 그 다른 변수와의 관계를 명시하여 해당 변수의 효과를 논해야 할 것이다. 식 (1)의 분석은 그런 면을 등한히 하고 있다.

4. 실증분석 예

다중회귀 모델을 분석하는 데 있어 제기되는 문제들을 실증분석으로 확인하기 위해 마케팅에서 많이 사용하고 있는 Fishbein모델을 이용하기로 하자. 여러 학자들은(e. g. , Liska, 1984) Fishbein모델에 깔려있는 가산적 가정에 의문을 제기하면서 행동에 대한 태도 (A_{act})와 주관적 규범 (SN)이 행동의도 (BI)에 가산적 효과 뿐만 아니라, 상호작용 효과도 함께 갖고 있다고 가설화하고 있다. 이 주장을 자동차 구매의 상황에 대한 연구에서 검증하였다. 본 연

구에서 분석 예는 Yi(1990)의 연구에서 얻은 자료를 이용하기로 한다. 118명의 응답자들을 대상으로 구매의도(BI)는 '차 구매의 가능성'이란 항목으로 11점 척도(+1,+11)를 사용해 측정되었고, 구매행동에 대한 태도(A_{act})는 "좋다-나쁘다"라는 7점척도(+1,+7)를 가지고 측정되었다. 또, 주관적 규범(SN)은 "나에게 중요한 대부분의 사람들이 내가 ____ 차를 구매해야 한다-구매해서는 안된다고 생각한다 라는 항목을 이용한 11점척도(+1,+11)를 사용해 측정되었다.

주효과와 상호작용 효과를 평가하기 위해A_{act}와 SN의 곱셈항을 도입하여 다음의 회귀식을 사용하였다.

$$BI = B_0 + B_1 A_{act} + B_2 SN + B_3 A_{act} SN$$

이 회귀분석에서 얻은 회귀계수와 각각의 표준오차는 (표 1)의 첫번째 열에 제시되었다. 표에서 알 수 있듯이, 어떠한 계수도 통계적으로 유의하지 않다는 것을 확인할 수 있다. 이는 Fishbein모델에서의 주장과는 반대로, A_{act}와 SN이 BI에 대해 상호작용효과는 물론 주효과도 가지지 못하고 있다는 결론을 유도한다. 그러나, 회귀방정식 전체는 0.001수준에서 매우

〈표 1〉 회귀계수에 대한 척도변환의 효과

	기본모델[a]	A_{act}척도를 변화시킨 경우[b]	SN척도를 변화시킨 경우[c]	A_{act}와 SN척도 모두를 변화시킨 경우[d]
태 도	0.74(0.92)[e]	0.74(0.92)	1.38*(0.59)	1.38**(0.59)
주관적 규범	1.30(0.79)	1.73**(0.23)	1.30(0.79)	1.73**(0.23)
상호작용효과	0.11(0.17)	0.11(0.17)	0.11(0.17)	0.11(0.17)
Y절편	2.5	5.5	10.4	15.9
R^2	0.44	0.44	0.44	0.44

[a] A_{act} (1,7), SN(1,11)
[b] A_{act} (−3,+3), SN(1,11)
[c] A_{act} (1,7), SN(−5,+5)
[d] A_{act} (−3,+3), SN(−5,+5)[e] 괄호안의 숫자는 표준오차임.
* 0.05수준에서 통계적으로 유의한 수치.
** 0.01수준에서 통계적으로 유의한 수치.

유의하다는 것에 주목해야 한다(R^2 = 0.44, F = 29.6). 이 결론은 분석에 있어 다중공선성이 존재한다는 것을 암시하고 있다. 전체 회귀등식의 유의도는 매우 높지만 각 추정치는 표준오차의 증가로 인해 유의적이지 못하기 때문이다.

다중공선성은 ① 독립변수들간의 일조 상관계수(pairwise correlation), ② 상관계수 행렬의 행렬식(determinant), det(R) , ③ 상관계수 행렬의 역행렬의 대각요소인 (diagonal elements) 분산증폭요인(variance inflation factors;VIF(j)), ④ 투입자료의 변화에 대한 추정치의 민감도를 나타내는 조건수(condition number)인 k(X)등의 네가지 지표들에 의해 평가된다(Belsley et al., 1980 ; Mansfield & Helms, 1982). 변수들간의 상관계수는 다중공선성을 확인하는데 중요한 자료가 된다. 그러나 변수들간의 상관관계만으로는 충분하지 않은데 그 이유는 다중공선성이란 것이 예측변수들간의 전반적인 상호의존도이기 때문이다. 한 상관관계가 매우 클때 다중공선성이 존재하지만 여러개의 상관계수들이 다중공선성을 야기시킬 수 있다. 예를 들어 한 상관계수가 0.8인 것 보다 0.6이상의 상관계수가 두개인 것이 더 심각한 다중공선성의 문제를 야기시킬 수 있는 것이다.(Rockwell,1975).

또, 예측변수들간의 상관계수행렬의 행렬식(determinant)도 다중공선성을 측정하는데 유용한 정보를 제공한다(Farrar and Glauber, 1967). 상관계수행렬의 행렬식은 예측변수 간의 상호의존도가 높아짐에 따라 작아진다. 모든 예측변수들이 독립적일때 행렬식은 1이며 완벽하게 다중공선성인 경우에는 0이다. 따라서 0에 가까운 행렬식은 다중공선성을 나타내는 것으로 해석할 수 있다. 조건수 k(X)는 상관계수행렬의 고유치(eigenvalue) 중 최대치와 최소치의 비율로써 계산되는데 투입자료의 작은 변화에 대해 추정치가 얼마나 민감한가를 반영한다. 구체적으로 독립변수간의 상관계수가 1에 가까울때, 상관계수행렬의 행렬식이 0에 가까울때, 분산증폭요인 VIF(j)가 클때, 조건수가 높을때 다중공선성의 문제가 존재하는 것이다.

(표 2)에서는 A_{act},SN, 그리고 곱셈항인 $A_{act}SN$간의 상관관계를 보이고 있다. 이때, SN과 곱셈항 사이의 상관관계는 0.92로 나타났으며 이는 높은 다중공선성이 존재함을 의미한다. 다른 지표를 통해서도 이와 유사한 결론을 얻어낼 수 있다(det(R) = 0.038, 최대 VIF(j) = 20, k(X) = 111.4).

척도 종속성을 검토하기 위해 독립변수의 코딩작업에서 세가지 변환을 시도하였다. 사용된 변환은 ① A_{act} 척도의 변환, ② SN 척도의 변환, ③ A_{act}척도와 SN척도의 동시 변환등이였다. 각 변환의 결과가 (표 1)의 제 2-4열에 제시되어 있다.

A_{act}척도를 (−5,+5)로 부여해 양극화 척도로 바꾼 첫번째 변환의 결과, *SN*의 회귀계수가 1.30에서 1.73으로 변동되었다. 게다가, 회귀계수의 통계적 유의도도 상승하였다($p<0.01$). 이와 유사한 방식으로 두번째 변환의 결과가 제시되었다. 즉, *SN*척도를 (−3,+3)으로 변환시켜 A_{act}의 계수가 0.74에서 1.38로 변동되었으며 그 통계적 유의도가 상승하였다($p<0.05$). A_{act}척도와 *SN*척도를 모두 변환시킨 세번째의 경우, 양 변수의 계수가 모두 변동하였다. A_{act}와 *SN*의 주효과가 기본모델에서는 유의하지 못하였으나 이제는 0.05수준에서 유의한 결과를 보이게 되었다.

〈표 2〉 변수들간의 상관관계

	Y^a	X_1^b	X_2^c
Y	1.00		
X_1	0.36	1.00	
X_2	0.64	0.32	1.00
X_1X_2	0.65	0.61	**0.92**
$(X_1-\overline{X_1})(X_2-\overline{X_2})$	0.17	0.06	0.21
$(X_1-f)(X_2-g)$	0.04	0.00	0.00

[a] Y = 행동의도(BI)
[b] X_1 = 행동에 대한 태도(A_{act})
[c] X_2 = 주관적 규범(SN)

이때, 주의해야 될 점은 위 세가지 변환의 경우 모두 상호작용 효과는 변화하지 않았다는 것이다. 즉, 모든 변환에서 상호작용 효과의 계수는 0.11(표준오차 = 0.17)의 수치를 보이고 있다. 또, 모든 모델의 결정계수 R^2는 동일하게 나타났다. 그러나, 주효과에 대한 결론은 각 변환마다 급격히 달라지게 된다는 것에 주목해야 된다.

위와 같은 결과들에 의해, 같은 자료를 가지고도 사용되는 코딩지침에 따라, 즉 채택된 척도점수에 따라 태도와 주관적 신념의 주효과가 다르게 해석될 수 있다는 것을 알 수 있다. 구간척도의 코딩지침에 대한 학자들사이의 합치된 견해가 없는 현실에서 이 사실은 매우 중요한 의미를 지닌다. 예를 들어, 7점척도는 1에서 7까지로 코딩될수도 있고, -3에서 +3으로 코딩될 수도 있는 것이다. 그러나 어떻게 변수를 코딩하느냐에 따라서 그 변수의 효과에 대한

결론이 달라지는 것을 보았다.

Ⅲ. 문제해결을 위한 대안적 수단들

두 변수(예를 들어, X_1과 Y)간의 관계가 다른 변수(X_2)의 값에 따라 변동되는 지를, 즉 상호작용을 평가하기 위한 여러 방법들이 제안되고 있다. 그 방법들중 몇가지는 본 연구에서 지적한 여러 문제점들을 해결하기 위한 수단으로 유용하다. 여기에서는 다음의 세가지 방법에 초점을 맞추기로 하자.

1. 소집단 분석 (subgroup analysis)

소집단 분석이란 X_2를 조정변수로 하여 전체표본을 여러 소집단으로 나누고 여러 소집단에서의 Y와 X_1의 차별적 관계를 직접적으로 검토하여 상호작용 효과를 확인하려는 절차이다. 즉, 전체 집단을 여러 소집단으로 나누어 각 소집단을 독립적으로 분석하겠다는 것이다. 이때, 분석의 초점은 한 소집단에서의 X_1의 회귀계수가 다른 소집단의 회귀계수와 동일한가 또는 다른가 여부에 놓여진다. 만약, 회귀계수가 유의하게 차이를 보인다면, 상호작용 효과가 존재한다는 결론을 내릴 수 있다. 이 분석은 식 (1)의 특수한 형태로 이해될 수 있다. 왜냐하면, 양분변수(예: 남,여)와 같이 집단을 구분하는 변수를 이용하면 소집단 분석을 할수 있기 때문이다.

소집단 분석은 이용하기가 쉽다는 이점을 가지나, 몇가지 단점을 가지고 있다. 첫째, 연속변수(예를 들어, X_2)의 값이 인위적으로 두가지 혹은 세가지 범주속으로 할당되는 경우, 측정정보의 손실이 발생할 수 있다(Cohen and Cohen, 1983). 예를 들어, 태도의 점수가 연속적으로 존재하는 데 호의적, 비호의적 집단의 두개로 나눌 때 각 집단내 개인적 차이가 무시되는 것이다. 태도를 1에서 7까지의 점수로 측정했을 때 1에서 4까지와 5에서 7까지를 각기 비호의적 집단과 호의적 집단으로 나눌 수 있다. 그럴 때 태도의 점수가 1이거나 4이거나 관계없이 동일하게 취급되어 비호의적 집단내 태도점수의 차이가 무시되는 것이다. 또, 마찬가

지로 호의적 집단내에서도 5,6,7의 점수에 대한 정보가 손실되는 셈이다. 이는 연속변수로 측정된 값이 범주로 나눌 때 불연속 변수로 바뀌며 발생하는 문제점이다.

둘째, 어떤 점수로 소집단을 나누는가에 따라 관찰된 관계가 급격히 변동될 수 있기에 소집단구성에 따라 애매모호한 결과가 제시될 수도 있다(Cronbach and Snow, 1977). 이 두번째 문제점은 전체 표본집단에서 구분될 수 있는 소집단의 수가 다양하기 때문에 야기된다고 평가된다. 특히 여러 소집단의 분류에 대한 사전적 이론을 가지지 못하는 경우에 그래서 소집단의 수가 임의로 정해질 수 있는 경우에 문제는 심각해진다. 예를 들어 X_2값에 따라 소집단으로 나누기로 한 경우에 저,고의 2개 집단으로 나눌지, 저,중,고의 3개집단으로 나눌지가 명확하지 않은 것이다. 또, 2개의 소집단으로 나누기로 한 경우에도 구체적으로 어떤 점수를 기준으로 소집단을 구분해야 하는지가 불분명하다. 또한, 소집단의 분류에 있어 어떤 변수(X_1 또는 X_2)를 사용해야 하는가에 대해서도 논란이 있다.

세째, 소집단 분류의 기초가 된 변수의 주효과에 대한 해석이 불분명하게 될 수 있다. 예를 들어, X_1을 기초로 소집단을 나눈 경우, 각 소집단내에서 X_1의 범위는 제한되기 때문에 X_1과 종속변수Y의 관계가 드러나지 않을 수 있다. 이는 결국, X_1의 주효과가 범위제한(range restriction)에 의해 유의하지 못한 것으로 판명될 수 있다는 것을 의미한다.

네째, 총 관측치의 수가 많지 않을 경우에는 소집단 분석을 하게되면, 각 소집단내의 관측치의 숫자가 크게 줄어들 수 있다. 따라서 회귀계수의 추정치가 불안정적일 수 있다.

2. 독립화 변환 (orthogonal centering)

Smith and Sasaki(1979)는 곱셈항을 그 구성요소들과 독립적으로(무관한 상태로) 만들도록 변수들을 변환해야 한다고 제안했다. 이 독립화 변환은 다음과 같은 등식으로 표현된다. 이때, 이 모델은 독립적으로 변환된 (orthogonally centered) 모델이라 불린다.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3(X_1 - f)(X_2 - g) + e \tag{3}$$

이때, 변수들을 상호독립적으로 바꿔주는 변환상수 f와 g는 다음과 같이 계산된다.

$$f = (S_{11}S_{23} - S_{12}S_{13}) / (S_{11}S_{22} - S_{12}^2),$$

$$g = (S_{13}S_{22} - S_{12}S_{23}) / (S_{11}S_{22} - S_{12}^2).$$

단, S_{ij}는 X_i와 X_j의 공분산을 의미하며, S_{ii}는 X_i의 분산을 의미한다. 그리고, X_3는 X_1 X_2를 나타낸다.

이 방법을 사용하면, 상호작용 효과를 나타내는 항과 X_1, X_2항과의 상관관계가 제거되어 다중공선성은 감소된다. 그러나, 계수 b_i의 추정은 여전히 척도의 변화에 따라 변동되는 단점을 그대로 갖게 된다. 즉, X_i와 X_j의 원점이 변동하면 이 계수들도 달라진다. 또, 주효과에 대한 해석도 여전히 곤란하다. 즉, 이 모델에서 X_1의 주효과는 다른 독립변수(X_2)가 곱셈항(X_1 X_2)과 독립변수사이의 상관관계를 없애주는 어떤 인위적인 값을 가질 때(즉, $X_2 = g$일 때)의 효과로 이해되는 것이다. 이런 단점들외에도, 이 독립화 변환은 계산이 복잡하다는 단점도 가지고 있다.

3. 평균변환 (mean centering)

다중회귀모델에서 사용되는 또 하나의 해결수단은 평균으로부터의 차이값을 이용하는 것이다. 이 방법은 X_1과 X_2의 값을 각각 평균과의 차이로 변형시키고 곱셈항도 이렇게 변형된 값으로 대체하는 것으로 구성된다. 다시 말하면, X_1과 X_2의 원점은 각 변수의 평균으로 이동시키는 것이다. 이 변환은 평균변환 (mean-centered) 모델로 표현되며 수식은 다음과 같다.

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + e \tag{4}$$

이때, $x_1 = X_1 - \overline{X_1}$,

$x_2 = X_2 - \overline{X_2}$,

평균변환은 상호작용 효과를 고려하는 회귀모델의 세가지 문제-척도 종속성, 다중공선성, 해석상의 문제등을 해결해 준다. 첫째, 평균변환모델은 기초모델에서 발생하는 척도 종속성의 문제를 해결해준다. 즉, 식 (4)의 회귀계수는 X_1과 X_2의 원점을 변화시킨다고 해서 변동되지 않는다. 예를 들어, X_1과 X_2척도의 원점이 $Z_1 = X_1 - c$, $Z_2 = X_2 - d$라는 식에 의해 변화된 경우, 평균변환을 하면 다음과 같은 결과가 제시된다.

$$z_1 = Z_1 - \overline{Z_1} = (X_1 - c) - (\overline{X_1} - c) = X_1 - \overline{X_1} = x_1,$$

$$z_2 = Z_2 - \overline{Z_2} = (X_2 - d) - (\overline{X_2} - d) = X_2 - \overline{X_2} = x_2.$$

즉, X_1과 X_2를 변환시킨 후에도 평균변환값인 x_1과 x_2는 변동되지 않는다. 그 결과, 식 (4)에서 평균변환 모델의 회귀계수는 변수의 척도변환에 영향을 받지 않게 된다. 결과적으로, 척도 종속성의 문제는 해결된 것이다.

둘째, 평균변환은 각 독립변수 X_i와 곱셈항 $X_1\ X_2$간의 상관관계를 감소시킨다(Cohen and Cohen, 1983). Tate(1984)에 의하면, 독립변수를 평균변환하면 (즉, 독립변수를 각각의 평균과의 차이로 표현하면) 다중공선성을 상당히 줄일 수 있으며 그 결과, 다중공선성과 관련된 문제들을 피할 수 있다고 한다.

이 사실을 다음과 같이 보일 수 있다. 변수에 평균변환을 적응시키면 그 변수들의 평균을 0으로 놓는 셈이다 : $E(x_1)=E(x_2)=0$. x_1과 x_2의 분포에 대해 어떤 가정이 없는 일반적인 경우에는 x_1x_2와 x_1, x_2간의 공분산은 다음과 같다(부록 2 참조).

$$Cov(x_1x_2, x_1)=E(x_1^2\,x_2)$$

$$Cov(x_1x_2, x_2)=E(x_1x_2^2)$$

만약, x_1과 x_2가 정규분포를 따르는 경우에는 홀수 모멘트가 0이므로 공분산은 다음과 같다.

$$Cov(x_1x_2, x_1)=Cov(x_1x_2, x_2)=0$$

즉, 정규분포를 가정하는 경우 평균변환을 하면 $x_1\ x_2$항과 x_1, x_2항간의 상관관계가 감소하게 되는 것이다. 또 X_1과 X_2의 분포에 대한 가정이 없는 일반적인 경우에도 다중공선성의 정도가 평균변환에 의해 종종 감소된다. (Neter, Wasserman and Kutner 1983). Aiken and West(1991)도 여러가지 평균변환자료와 원자료의 비교를 통해 이 점을 보였다. 그러나 물론 평균변환이 다중공선성의 문제를 해결시켜주지 못하는 경우도 있을 수 있겠다. 보다 구체적으로 보면 $Cov(X_1X_2, X_1)$과 $Cov(x_1x_2, x_1)$의 차이가 $V(X_1)E(X_2)+Cov(X_1, X_2)E(X_1)$이다. (부록1과2참조).

그러므로 $E(X_1)$, $E(X_2)$, $Cov(X_1, X_2)$의 부호와 크기에 따라서 평균변환에 의해 다중공선성이 감소될 수 있는가의 여부나 정도가 달라질 것이다.[1)]
일반적으로 모집단에 있어서 변수간에 실제로 관계가 존재하기 때문에 생기는 다중공선성

1) 이 점을 지적해 주신 사독위원에게 깊은 감사를 드립니다.

은 필수적인 불량조건(essential ill-conditioning)이라고 한다. 반면에 상호작용변수인 곱셈항때문에 생기는 다중공선성은 불필요한 불량조건 (nonessential ill-conditioning)이라고 한다 (Marquardt, 1980). 이 문제는 척도화에 따른 부작용으로 평균변환에 의해 줄일 수 있다.

세째, 평균변환은 X_1과 X_2의 주효과를 해석하는 데도 도움을 준다. 즉, 한 독립변수 X_1의 주효과는 다른 변수 X_2의 모든 가능한 값에 걸친 X_1의 평균효과로 해석된다. 만약, 식 (4)을 가지고 X_1에 대한 편미분을 구하면 다음과 같다.

$$\partial Y / \partial X_1 = a_1 + a_3 x_2 = a_1 + a_3 (X_2 - \overline{X_2})$$

이 계산결과, X_1의 주효과인 b_1은 다른 독립변수 X_2가 그 자신의 평균과 동일한 값을 가질 때, X_1의 효과라는 것을 알 수 있다.

위와 같은 주효과의 해석은 여러 학자들에 의해 제안되고 있는 것과 유사하다. 상호작용 효과가 있을 때, 주효과를 추정하는 것은 타당하지만 그 때 주효과의 해석은 상호작용이 없을 때와는 달라진다(Overall et al., 1981). 상호작용을 고려하지 않는 모델에서는 ($Y = B_0 + B_1X_1 + B_2X_2$) 한 변수의 주효과가 다른 변수의 값에 관계없이 일정하게 나타난다. 그러나, 상호작용을 고려하는 모델에서는 주효과가 다른 변수의 값의 변동에 따라 변화하게 된다. 그런 경우, 주효과는 다른 독립변수의 모든 가능한 값에 걸친 해당 변수의 평균적 효과로 해석되어야 한다. 이런 의미에서의 주효과는 다른 독립변수의 평균값에 있어서 해당 변수의 효과로 추정될 수 있다(Cohen and Cohen, 1983 ; Lane, 1981).

본 연구에서 제안된 평균변환 모델은 위와 같은 주효과를 추정할 수 있도록 해준다. 게다가, 주효과가 다른 변수와의 관계를 기초로 해석된다는 이점을 얻게 된다.

평균변환 모델은 추가적으로 다음과 같은 이점을 가진다. 첫째, 식 (1)에서의 계수 B_i에 대한 추정이 식 (4)에서의 계수 a_i와 선형적 관계를 가지기에, a_i를 알면 쉽게 B_i도 계산할 수 있게 된다.

둘째, 식 (1)과 식 (4)는 수학적으로 동일하기 때문에, 양자는 자료에 대한 동일한 부합도를 보인다. 평균변환은 척도 종속성, 다중공선성을 극복하면서 주효과에 대한 명확한 해석도 가능케 해주는 동시에, R^2의 값도 변화시키지 않는다.

세째, 평균변환 모델은 두가지 독립변수에만 적용되는 것이 아니라, 그 이상의 독립변수가

사용되는 경우에 대해서도 쉽게 일반화될 수 있다.

네째, 상호작용 효과에 대한 해석이 명확해졌다. 식 (4)에서 알 수 있듯이, 상호작용 효과는 독립변수가 평균과의 차이를 크게 가질수록 커진다. 이때, 두 독립변수가 평균으로부터 같은 방향으로 차이가 난다면 상호작용 효과가 陽(+)이 되지만, 반대로 서로 다른 방향으로 차이가 나는 경우 (한 변수는 평균보다 크고, 다른 변수는 평균보다 작은 경우), 상호작용 효과는 陰(−)이 될 것이다. 이렇게 볼 때, 상호작용 효과는 일관성 효과 (consistency effect)라 해석될 수 있다. 결론적으로, 위와 같은 이점을 가지고 평균변환이 주효과와 상호작용 효과의 평가에 가장 유용한 수단이라는 것을 확인할 수 있다.

그러나, 평균변환이 항상 필요한 것은 아니다. 즉, 한 독립변수가 집단의 구성을 정의하는 가변수(dummy variable)인 경우, 기초모델의 상호작용 효과를 나타내는 항의 계수는 0으로 코딩되는 기초집단과 해당 집단간 기울기의 차이를 나타내게 된다. 연속변수의 계수는 기초집단의 해당 변수의 기울기를 의미한다. 예를 들어, 가변수인 X_2의 값이 남성에게는 0으로 여성에게는 1로 정해진다면 X_1 X_2항의 계수는 남성과 여성에 있어서 기울기의 차이를 나타낸다. 그리고 연속변수인 X_1의 계수는 기초집단인 남성에 있어서 X_1의 기울기를 나타낸다. 게다가, 두 독립변수가 팩토리얼 디자인과 같은 실험연구에서의 예측변수로 사용되는 경우, 기초모델에서 나타나는 곱셈항의 계수를 해석하기는 쉽다. 즉, 그 계수는 분산분석(ANOVA)에서의 상호작용 효과와 동일하게 되는 것이다. 그런 경우에, 기초모델만을 가지고도 충분한 해석을 할 수 있다. 이와 같이 현실적으로 평균변환이 필요없는 경우도 존재하게 된다.

4. 실증분석 예

앞에서 사용한 자료에 평균변환을 적용해 보기로 하자. 즉, Fishbein모델에서의 A_{act}와 SN의 효과를 평가하기 위해 각 변수 값에 평균과의 차이를 도입해 보자. (표 2)에서 제시되는 상관관계 행렬은 평균변환 기법이 예측변수간의 상관관계를 상당히 축소시켜 준다는 것을 보여준다. 예를 들어, 곱셈항과 A_{act}의 상관계수가 0.61에서 0.06으로 낮아졌고 곱셈항과 SN의 상관계수도 0.92에서 0.21로 낮아졌다. 다른 지표들도 다중공선성이 감소하였다는 것을 보여주고 있다. 즉, 행렬식 det(R)=0.858, 최대 분산증폭요인 VIF(j)=1.2, 조건수 k(X)=9.9이

다. 또한, 평균변환에 의해 A_{act}와 SN의 계수가 갖는 표준오차도 상당히 감소하였다. 예를 들어, A_{act}의 계수에 대한 표준오차는 0.92에서 0.51로 감소하였고, SN의 계수에 대한 표준오차는 0.79에서 0.23으로 감소하였다. 이와 같은 표준오차의 감소는 다중공선성의 감소로 발생하는 결과이다. 전반적으로 볼 때, 평균변환에 의해 전체적으로 다중공선성이 상당히 해결되었다는 것을 알 수 있다.

다음에서는 척도 종속성의 문제를 검토해보자. 평균으로부터의 편차는 척도의 원점변환과는 무관하다는 것을 앞에서도 확인하였다. 즉, 회귀계수와 그에 따른 표준오차가 선형변환에 아무런 영향을 받지 않으므로 척도의 원점변환에 의해 주효과와 상호작용효과에 대한 결론이 영향받지 않는다.

비교하기 위하여 독립화 변환도 적용해 보았다. (표 2)에서 알 수 있듯이, 곱셈항과 그 구성변수간의 상관관계는 독립화 변환에 의해 사라졌음을 알 수 있다. 그러나, 다중공선성의 지표들은 평균변환과 매우 유사하게 제시된다 ; det(R)=0.898, 최대 VIF(j)=1.1, k(X)=9.7.독립화 변환에 의한 계수와 표준오차가 (표 3)에 제시되어 있다. 표에서 알 수 있듯이, 해당 계수와 표준오차가 평균변환의 결과와 유사하게 나타나고 있다. 예를 들어, A_{act}의 계수는 독립화 변환에서 1.21으로 나타났으며, 평균변환 모델에서는 이와 유사하게 1.20으로 제시되었다.

또, 소집단 분석을 같은 자료에 적용해 보았다. 구체적으로 A_{act}를 기초로 표본을 소집단으로 나누어 소집단 분석을 시행하였다. 즉, 중앙값을 기초로 '낮은 태도 집단'과 '높은 태도 집단'으로 구분하였다. 각각의 집단에 대해 회귀모델을 추정하였고 그 결과는 (표 4)에 제시되어 있다.

표에서 알 수 있듯이, 태도가 중앙값보다 낮은 집단에서의 A_{act}의 계수는 1.27이고 그 표준오차는 0.86이다. 이 계수는 0.10수준에서는 통계적으로 유의하지 못하다. 반면에, SN은 1.64의 계수를 가지고 유의한 효과를 가지고 있다. 태도가 높은 집단에서도 이와 유사한 결과가 제시되고 있다. 즉, SN은 BI에 유의한 효과를 가지는 반면, A_{act}는 그렇지 못한 것으로 확인되었다. 이와 같이, 집단을 분리해 분석을 시행하는 경우 곱셈항이 사용되지 않기에 다중공선성은 문제가 되지 않는다. 다중공선성에 대한 측정은 이와 같은 주장을 뒷받침해주고 있다. 즉, 태도가 낮은 집단에서는 최대 일조 상관계수=0.53, det(R)=0.695, 최대 VIF(j)=1.0, k(X)=4.0의 결과가 제시되었고, 태도가 높은 집단에서는 최대 일조 상관계수=0.72, det

(R)=0.410, 최대 VIF(j)=1.1, k(X)=4.5의 결과가 제시되었다. 따라서, 독립화변환을 사용하기 위해 요구되는 계산의 복잡성과 주효과 해석상의 곤란한 정도에 비해 얻는 이점이 그다지 많지 않다는 것을 확인할 수 있다.

〈표 3〉 평균변환 모델과 독립화 변환 모델의 추정결과

	평균변환 모델[a]		독립화 변환 모델[b]	
	계수	표준오차	계수	표준오차
태도	1.20*	0.51	1.21*	0.51
주관적 규범	1.74**	0.23	1.77**	0.22
상호작용	0.11	0.17	0.11	0.17
Y절편	13.2		13.5	
R^2	0.44		0.44	

[a] A_{act}의 평균 4.12와 SN의 평균 4.40으로 변환한 모델.
[b] SN(4.38, 4.39)에 대한 독립화변환모델
* 0.05수준에서 통계적으로 유의한 수치
** 0.01수준에서 통계적으로 유의한 수치

〈표 4〉 소집단 분석의 결과

	낮은 태도 집단[a]		높은 태도 집단[b]	
	계수	표준오차	계수	표준오차
태도	1.27	0.86	2.70	1.80
주관적 규범	1.64**	0.31	1.94**	0.34
Y절편	1.05		−8.65	
R^2	0.30		0.54	

[a] $A_{act} \leq 4$.
[b] $A_{act} > 4$.
**0.01수준에서 통계적으로 유의한 수치

그러나, 소집단 분석에 있어 주효과와 상호작용 효과의 해석은 불분명하다. 특히, A_{act}의 주

효과는 해석에 문제가 있는 것으로 확인되었다. 집단구분의 기초를 A_{act}의 점수에 놓았기에, 각 집단내에서 A_{act}의 범위가 제한되므로 A_{act}와 BI의 상관관계가 비유의적인 것으로 추정될 것이다. 실제로 분석의 결과, A_{act}의 계수는 각 집단에서 유의하지 못한 것으로 확인되었다. 평균변환이나 독립화 변환의 결과와 소집단 분석의 결과를 비교할 때, A_{act}의 주효과가 유의하게 나타나지 못한 것은 범위제한의 결과로 야기되는 추정상의 오류로 볼 수 있다. 게다가, 소집단 분석에서는 상호작용 효과에 대한 추정치를 직접 얻기 불가능하며 그에 따라, 상호작용 효과의 유의성 여부도 확인이 불가능하다.

상호작용 효과의 유의성을 검토하기 위해서는 식에 가변수(dummy variable)를 도입해야 한다. 즉, A_{act}의 두가지 수준에 대해 가변수를 만들고, 이 가변수와 SN과의 곱셈항을 모델에 도입한다는 의미이다. 즉, 중앙값보다 낮은 경우는 0을, 중앙값이 보다 높은 경우는 1을 부여하고 이 가변수와 SN과의 곱셈항을 도입한다는 의미로 당연히 가변수간의 곱셈항도 도입된다. 그리고 나서, 곱셈항의 계수가 A_{act}와 SN의 상호작용 효과로 해석된다. 이때, 통계적 유의도도 함께 검증된다. 가변수를 모델에 도입한 결과, R^2가 0.44로 제시되었다. 곱셈항의 계수는 0.39 (표준오차는 0.46)로 제시되었으며, 해당 상호작용 효과는 0.40수준에서 유의하지 못한 것으로 판명되었다. 물론, 가변수를 도입함으로써 상호작용 효과를 검토할 수 있었지만 분석에 다중공선성의 문제가 야기되었다. 이는 곱셈항이 일반적으로 A_{act}와 SN에 높은 상관관계를 가지기 때문이다. 분석결과, 높은 다중공선성이 존재하는 것으로 확인되었다(최대 일조상관관계 = 0.85, det(R) = 0.036, 최대 VIF(j) = 5.9, k(X) = 61.9).

요약해 보면, 평균변환이 다중공선성이란 문제해결의 가장 만족스로운 도구임을 알 수 있다. 독립화 변환을 이용하여 얻게 되는 다중공선성의 추가적 감소는 미미하다. 또, 독립화 변환에 사용되는 상수 f와 g는 X_1과 X_2의 평균과 매우 유사한 값을 가진다 ;

$$f = 4.38,\ g = 4.39,\ \overline{X_1} = 4.12,\ \overline{X_2} = 4.40$$

이는 Smith and Sasaki(1979)의 예측과 유사한 결과이다. 물론, 소집단 분석에 의해 다중공선성이 감소되나, 주효과와 상호작용 효과에 대한 해석과 상호작용 효과의 검증은 불분명해진다. 대조적으로, 가변수를 도입하여 소집단 분석을 행하는 경우, 상호작용 효과를 검증할 수 있으나, 다중공선성의 문제가 제기되는 것이다. 결론적으로, 평균변환은 주효과와 상호작용 효과를 평가하는데 있어 제기되는 세가지 문제를 해결하는 가장 유용한 수단임을 알 수 있다.

본 연구에서 살펴본 예의 경우, 상호작용 효과가 비유의적이었다. 그렇기 때문에 비유의적인 상호작용변수(X_1 X_2)를 회귀등식에서 제거해야 된다고 생각할 수도 있으나 비유의적이라고 해서 그 항목을 삭제하는 것에는 반대의견이 많다. 비유의적이라고 해서 어떤 변수를 없애는 것은 자료에 따라서 이용하는 통계모델을 바꾸는 것으로 이론에 기초한 모델을 선정과는 거리가 멀다. 그래서, 비유의적인 변수를 남겨두고 그 비유의도를 보고하는 것이 가치가 있다고 하겠다. 비유의적인 결과가 보고될 때 현상에 대한 지식이 보다 더 효율적으로 축적될 수 있다.

상호작용 효과가 비유의적이라고 해서 제거하게 되면 회귀계수의 추정치가 편의(bias)를 가질 수 있다. 특히, 연구설계가 통계적 검증력이 낮은 경우에 현실적으로 존재하는 효과를 발견하지 못할 수도 있다. 이와 같이 관련있는 변수를 생략하면 심각한 문제가 야기된다. 그리고, 추정치와 추정치의 분산도 편의성을 띄게 되며 통계적 추론이 부정확해질 수 있다. 반면에 무관한 변수가 있는 경우에는 회귀계수가 여전히 불편추정치가 된다. 그러므로, 자료와 자유도가 허용하는 한 변수를 삭제하기 보다는 포함하는 쪽으로 하는 것이 안전할 것이다(Johnston, 1972). 그러므로, 이론에 의해 상호작용을 설정한 본 연구에서는 비유의적인 상호작용 효과를 포함한 회귀분석의 결과를 제시하였다.

그러나 때로는 유의하지 않은 상호작용항을 회귀식에서 제거하는 것이 필요할 것이다. 회귀분석의 목적이 어떤 현상에 대한 이해를 증진시키는데에 있다기보다는 앞으로 벌어질 일에 대한 예측을 하는 것일 수도 있다. 이런 경우에 예측능력을 향상시키기 위해서는 상호작용항을 회귀식에서 제거하는 것이 도움이 된다. 회귀식의 예측능력을 나타내는 MSE(Mean Squared Error)가 상호작용항을 제거함으로써 개선될 수 있기 때문이다(Hagerty, 1986). 따라서 회귀분석의 목적이 예측인 경우에는 유의하지 않은 상호작용항을 회귀식에서 제거하여야 할 것이다. 이 경우에는 주효과의 평가에 관련된 문제들이 발생하지 않게 될 것이다.

Ⅳ. 결 론

지금까지 상호작용 효과가 존재하는 회귀분석에서 구간척도 측정치를 사용할 때 제기되는

문제점들을 살펴보았다. 과거 문헌을 검토해 보면, 상호작용 효과에 대한 분석에 대해서는 많은 연구가 진행된 반면, 상호작용항이 포함된 회귀식의 주효과 평가의 문제에 대해서는 큰 관심이 없었다. 본 연구는 상호작용항이 포함된 회귀식의 주효과 평가에 제기될 수 있는 여러 이슈들을 다루고 있다. 전통적인 분석은 척도 종속성, 다중공선성, 애매모호한 해석등의 문제점들을 갖고 있다. 본 연구에서는 일반적으로 다중공선성의 해결수단으로 제안되는 평균변환이 다른 문제들의 해결에도 유용하다는 것을 보였다. 즉, 평균변환을 이용하면 척도의 원점변환에 대해 결론이 아무런 영향을 받지 않는 동시에 다중공선성을 감소시킬 수 있고 또, 주효과와 상호작용 효과의 추정치가 명확하게 해석된다는 장점을 얻을 수 있다.

참 고 문 헌

1. 채 서일, 사회과학 조사방법론, 학현사, 서울, 1993.
2. Aiken, L. S. and S. G. West (1991), Multiple Regression : Testing and Interpreting Interactions, Newbury Park, CA : Sage
3. Allison, Paul D. (1977), "Testing for Interaction in Multiple Regression," American Journal of Sociology, 83 (1), 144-153.
4. Althauser, Robert P. (1971), "Multicollinearity and Non−additive Models," in Causal Models in the Social Sciences, H. M. Blalock, Jr., ed., Chicago: Aldine−Atherton, 433-472.
5. Bagozzi, Richard P. (1984), "Expectancy−Value Attitude Models: An Analysis of Critical Measurement Issues," International Journal of Research in Marketing, 1, 295-310.
6. Bagozzi, Richard P., Hans Baumgartner, and Youjae Yi (1992), "State versus Action Orientation and the Theory of Reasoned Action: An Application to Coupon Usage," Journal of Consumer Research 18(March), 505-518.
7. Belsley, D. A., E. Kuh, & R. E. Welsh (1980), Regression Diagnostics: Identifying Influential Data and Source of Collinearity, New York: Wiley.

8. Bennett, Peter D. and Gilbert D. Harrell (1975), "The Role of Confidence in Understanding and Predicting Buyers' Attitudes and Purchase Intentions," Journal of Consumer Research, 2 (September), 110-117.
9. Cohen, Jacob (1978), "Partialed Products Are Interactions: Partialed Powers Are Curve Components," Psychological Bulletin, 85 (3), 858-866.
10. Cohen, Jacob and Patricia Cohen (1983), Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, New York: John Wiley and Sons, Inc.
11. Cronbach, Lee J. and Richard E. Snow (1977), Aptitudes and Instructional Methods: A Handbook for Research on Interactions, New York: Irvington.
12. Draper, N. R. and H. Smith (1981), Applied Regression Analysis, 2nd ed., New York: John Wiley and Sons, Inc.
13. Farrar, D. E. and R. R. Glauber (1967), "Multicollinearity in Regression Analysis: The Problems Revisited," Review of Economics and Statistics, 49 (February), 92-107.
14. Hagerty, Michael (1986), "The Cost of Simplifying Preference Models", Marketing Science, 5(Fall), 298-319.
15. Haitovsky, Yoel (1969), Multicollinearity in Regression Analysis: A Comment," Review of Economics and Statistics, 51 (November), 486-489.
16. Holbrook, Morris B. (1977), "Comparing Multiattribute Attitude Models by Optimal Scaling," Journal of Consumer Research, 4 (December), 165-171.
17. Hornik, Jacob (1984), "Subjective vs. Objective Time Measures: A Note on the Perception of Time in Consumer Research, Journal of Consumer Research, 11 (June), 615-618.
18. Johnston, J. (1972), Econometric Methods, New York: McGraw-Hill.
19. Lane, David M. (1981), "Testing Main Effects of Continuous Variables in Nonadditive Models," Multivariate Behavioral Research, 16 (October), 499-509.
20. Laroche, Michel and John A. Howard (1980), "Nonlinear Relations in a Complex Model of Buyer Behavior," Journal of Consumer Research, 6 (March), 377-388.

21. Lilien, Gary L. and Philip Kotler (1983), Marketing Decision Making: A Model-Building Approach, New York: Harper & Row.

22. Louviere, Jordan J. and George Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocative Experiments: An Approach Based on Aggregate Data," Journal of Consumer Research, 20 (November), 35-367.

23. Mansfield, E. R. and B. P. Helms (1982), "Detecting Multicollinearity," The American Statistician, 36(3), 158-160.

24. Marquardt, D. W. (1980), "You Should Standardize the Predictor Variables in Your Regression Models", Journal of the American Statistical Association, 75, 87-91.

25. Neter, John, William Wasserman and Michael Kutner (1983), Applied Linear Regression Models, Homewood, IL: Irwin.

26. Overall, John E., Dennis M. Lee, and Chris W. Hornick (1981), "Comparison of Two Strategies for Analysis of Variance in Nonorthogonal Designs," Psychological Bulletin, 90 (2), 367-375.

27. Punj, Girish N. and David W. Steward (1983), "An Interaction Framework of Consumer Decision Making," Journal of Consumer Research, 10 (September), 181-196.

28. Rockwell, Richard C. (1975), "Assessment of Multicollinearity: The Haitovsky Test of the Determinant," Sociological Methods and Research, 3 (February), 308-320.

29. Rogosa, David (1980), "Comparing Nonparallel Regression Lines," Psychological Bulletin, 88 (2), 307-321.

30. Ryan, Michael J. and E. H. Bonfield (1975), "The Fishbein Extended Model and Consumer Behavior," Journal of Consumer Research, 2 (September), 119-136.

31. Schmidt, Frank L. and Terry C. Wilson (1975), "Expectancy Value Models of Attitude Measurement: A Measurement Problem," Journal of Marketing Research, 12 (August), 366-368.

32. Sharma, Subhash, Richard M. Durand, and Oded Gur-Arie (1981), "Identification and Analysis of Moderator Variables," Journal of Marketing Research, 18 (August), 291-300.

33. Smith, Kent W. and M. S. Sasaki (1979), "Decreasing Multicollinearity: A Method for Models with Multiplicative Functions," Sociological Methods and Research, 8 (August), 35-56.

34. Southwood, Kenneth E. (1978), "Substantive Theory and Statistical Interaction: Five Models," American Journal of Sociology, 83 (5), 1154-1203.

35. Tate, Richard L (1984), "Limitations of Centering for Interactive Models," Sociological Methods and Research, 13 (November), 251-271.

36. Yi, Youjae (1989), "On the Evaluation of Main Effects in Multiplicative Regression Models," Journal of the Market Research Society, 31(1), 133-138.

37. Yi, Youjae (1990), "The Indirect Effects of Advertisements Designed to Change Product Attribute Beliefs," Psychology and Marketing, 7(1), 47-63.

38. Zedeck, Sheldon (1971), "Problems with the Use of 'Moderator' Variables," Psychological Bulletin, 76 (4), 295-310.

부 록 1

$x_1 = X_1 - E(X_1), \qquad x_2 = X_2 - E(X_2)$

$V(X_1) = E(x_1^2), \qquad V(X_2) = E(x_2^2)$

$Cov(X_1, X_2) = E(x_1 x_2)$

X_1과 X_2의 곱셈항인 X_1X_2에 대해 살펴보기로 한다.

$X_1X_2 = [\, x_1 + E(X_1)\,][x_2 + E(X_2)]$

$X_1X_2 = x_1x_2 + x_1\, E(X_2) + x_2\, E(X_1) + E(X_1)E(X_2)$ ……(1)

따라서 X_1X_2의 평균은 다음과 같다.

$E(X_1X_2) = E(x_1x_2) + E(x_1)E(X_2) + E(X_1)\, E(x_2) + E(X_1)E(X_2)$

$E(x_1) = E(x_2) = 0$이므로

$E(X_1X_2) = E(x_1x_2) + E(X_1)E(X_2)$ ……(2)

X_1X_2항과 X_1의 공분산을 구하면 다음과 같다.

$$
\begin{aligned}
Cov(X_1X_2, X_1) &= E\{[X_1X_2 - E(X_1X_2)][X_1 - E(X_1)]\} \\
&= E\{[x_1x_2 + x_2E(X_1) + E(X_2)x_1]x_1\} \\
&= E(x_1^2x_2) + E(X_1)E(x_1x_2) + E(x_1^2)E(X_2) \\
&= E(x_1^2x_2) + E(X_1)Cov(X_1, X_2) + V(X_1)E(X_2) \quad \text{……(3)}
\end{aligned}
$$

마찬가지로 X_1X_2항과 X_2의 공분산을 구할 수 있다.

$Cov(X_1X_2, X_2) = E(x_2^2x_1) + E(X_2)Cov(X_1, X_2) + V(X_2)E(X_1)$ ……(4)

부 록 2

$$Cov(x_1x_2, x_1) = E(x_1^2x_2) + E(x_1)Cov(x_1, x_2) + V(x_1)E(x_2)$$
$$= E(x_1^2x_2)$$

(식(3) 참조와 $E(x_1) = E(x_2) = 0$)

$$Cov(x_1x_2, x_2) = E(x_1x_2^2) + E(x_2)Cov(x_1, x_2) + V(x_2)E(x_1)$$
$$= E(x_1x_2^2)$$