원전주변 주민의 최대개인섭취량 계산 방법에 관한 연구

조한석, 김주한, 박봉서, 이갑복*
충남대학교 정보통계학과, 대전광역시 유성구 궁동 110번지
*한전전력연구원, 대전광역시 유성구 문지동 487-0번지
hnsokjo@cnu.ac.kr

현재까지 국내 원자력관련시설 주변 주민의 방사선량 평가에 사용되고 있는 음식물 섭취량은 1987년 고리 원자력발전소를 대상으로 현장 조사한 섭취량 값을 전체 원자력관련부지에 대해 동일하게 적용하고 있다. 당시 설정된 최대개인섭취량의 값은 조사대상 800가구 중 섭취량이 많은 상위 10가구의 식품군별 평균 섭취량으로 산출되어 다소 보수적인 수치가 적용되고 있는 문제점이 있으며, 이러한 음식물 섭취량의 자료는 매 5년이내 조사를 실시한 후 정기적으로 새로운 값을 대체하도록 되어 있다. 그러나 인력과 예산상의 문제로 부지별 조사가 정기적으로 이루어지지 않고 있는 실정이다. 이러한 문제를 해결하기 위하여 보건복지부에서 국민건강증진법에 의거하여 3년을 주기로 조사되는 전국 규모의 국민건강·영양조사 자료를 활용하여 주기적인 음식물 섭취량 자료를 갱신하는 방안이 모색되었다. 또한 국제원자력기구(IAEA)와 선진 국가에서는 해당 국가의 특성에 맞는 음식물의 범주를 설정하여 각 식품군의 최대섭취량을 0.95 분위수 혹은 0.975 분위수를 활용하도록 권고하거나 사용하고 있으며, 우리나라도 이러한 분위수를 활용하여 합리적인 섭취량 값을 도출하는 방안이 요구되었다.

현재 사용 중인 주민피폭선량 평가방법에 맞는 식품군별 섭취량을 정의하여야 하며 이에 타당한 식품군별 섭취량을 추정할 때의 문제점을 살펴보면 다음과 같다. 국민건강·영양조사에서는 각 식품별로 일일섭취량이 조사되었으며, 방사선학적 오염경로의 특성에 맞게 몇 개의 식품군으로 분류할 수 있다. 이때 각 식품군의 섭취량을 $X=(X_1, X_2, \cdots, X_k)$, 각 식품군의 p 분위수를 $\xi_{1p}, \xi_{2p}, \cdots, \xi_{kp}$ 라 할 때, 전체 식품 섭취량의 q 분위수를 ξ_{Yq}라 하면, $p=q$ 일 때 각각 식품군 섭취량의 p 분위수들의 합은 전체 식품섭취량의 q 분위수보다 크게 되어 합리적인 섭취량으로 인정하기 어렵게 된다. 이러한 현상은 실제 2001년도에 조사된 4계절 국민영양조사 섭취량 자료 뿐만 아니라 대학수학능력시험의 표준화 점수에서도 찾아볼 수 있다.

이러한 문제점을 해결하기 위하여 다변량자료에서 각 변수들에 대한 합리적인 분위수 값을 추정할 수 있는 세 가지 방법을 제안한다.

첫째, 식품섭취량 총합의 분위수가 주어질 때 평균과의 차이를 각 변수들의 평균에 적절하게 더하여 분배하는 방법이다. 즉, 최대개인섭취량을 $Y=\xi_q$라고 정의하고, 방사선평가에 사용하는 각 식품군 섭취량을 $Y=\xi_q$일 때의 조건부평균으로 사용하는 방법이다. 다시 말해서 식품군의 평균 음식물 섭취량과 편차를 이용하는 방법으로 총섭취량의 q 분위수와 평균값의 차이에 식품군별 섭취량의 공분산을 고려한 가중치를 곱하여 각 식품군 섭취량의 평균에 더하여 주는 방법이라고 할 수 있다. 이 방법으로부터 얻은 각 식품군의 섭취량은, 총 섭취량이 주어졌을 때 각 식품군 섭취량의 조건부분포의 평균이고, 분산-공분산 행렬을 이용한 가중치를 적용하여 변수들의 크기와 상관관계를 모두 고려하여 최대개인섭취량을 결정하는 방법이다. 이 방법은 각 식품군들의 섭취 양상을 반영한다는 장점이 있지만, 자료가 다변량 정규분포를 가정할 수 없는 경우에 적용이 어려운 단점이 있다.

둘째, 다변량 변수들의 합에 대한 특정 분위수 ξ_q와 각 다변량 변수들의 p 분위수 값의 합 $(\eta_1+\eta_2+\cdots+\eta_k)$이 같게 되는 $(\eta_1, \eta_2, \cdots, \eta_k)$를 찾을 수 있다. 이러한 경우 다변량 변수들의 분산과 상관관계를 고려하여 p와 q의 관계를 정리할 수 있다. 그러나 이 방법은 계산이 복

잡하고 자료의 특이값과 이상치에 영향을 받을 수 있는 단점이 있다.

세 번째 방법은 Chaudhuri(1996)가 소개한 다변량 자료에서의 기하 분위수(Geometric Quantile)를 활용하는 방법으로 Newton-Raphson 방법을 사용하여 기하분위수를 계산할 수 있다. 앞서 거론된 방법보다는 이론적인 객관성이 확보되는 방법이라고 할 수 있다. 그러나 이 방법을 적용한 결과 앞서 거론된 방법들의 적용결과와 비교할 때 표준편차가 큰 변수는 상대적으로 기하 분위수가 감소하였으며, 표준편차가 작은 변수는 상대적으로 기하 분위수가 증가되었다는 것을 알 수 있었다. 또한 자료가 극값에 영향을 받아 치우쳐진 분포의 형태인 경우 이를 로그변환하여 기하 분위수를 계산한 후, 변환 전의 값으로 환산하였을 때, 실제 자료의 범위를 벗어나 상당히 큰 값이 계산 되었다. 즉, 다변량 기하 분위수는 $\|u\|$의 값이 1에 근접하면서 표준편차가 큰 변수는 상대적으로 낮게, 표준편차가 작은 변수는 상대적으로 높게 계산되고, 자료를 로그변환한 후 반환된 값과의 차이가 크며, 원자료의 범위에서 벗어난 분위수 값이 계산되는 불합리성을 내포하고 있다. 기하 분위수는 유클리드 거리 개념을 사용하고 있으며, 따라서 통계분야에서 주로 고려되는 확률적인 측도에 바로 적용하기가 부적당하다. 이러한 이유는 각 다변량 변수가 갖는 서로 다른 변이성을 반영하고 있지 못한 것에서 기인하는 것이라 할 수 있다.

네 번째 방법은 기하분위수에 대한 대안으로 제안하는 것으로써 표준화 기하 분위수는 앞서 거론된 기하 분위수의 불합리성을 보완하기 위하여 각 변수를 그의 표준편차로 나누어 척도를 변환시킨 값에 대하여 유클리드 거리 개념을 적용한 후 다시 이전의 값으로 변환하는 것이다. 계산결과 표준화 기하분위수를 사용하는 것이 대부분 자료의 범위 안에서 현실적인 값으로 계산되어지며, 로그변환을 취한 값으로 계산한 후 결과값을 역변환 하더라도 그 값의 차이가 그리 커지지 않는 다는 것을 알 수 있다. 그러므로 자료가 반드시 다변량 정규분포를 따르지 않더라도 표준화 기하 분위수를 이용하는 것이 합리적인 분위수를 계산할 수 있었다.

이상과 같이 제안된 방법에 따라 국민영양조사 자료에 대한 섭취량의 계산결과 중 17세이상의 성인에 대한 자료를 <표. 1>에 제시하였으며, 1987년부터 최근까지 방사선평가에 사용된 기존 섭취량값과 함께 제시하였다. 총섭취량의 0.95 분위수 값을 최대개인섭취량으로 결정할 때, 이에 대한 각 11개 식품군에 대한 섭취량으로 각 방법별로 계산한 값보다 적게 먹는 사람의 수에 대한 확률도 계산하였으며, 기존 식품군 섭취량 보다 작은 확률이 가장 크게 나타났고, 표준화 기하 분위수 값보다 작은 확률이 가장 작게 나타났다.

<표 3. 11> 성인연령군 최대개인음식물 섭취량

식품군	기존섭취량	첫 번째 방법	두번째 방법	세번째 방법	네번째 방법
곡류	516.5	434.2	462.7	427.8	469.2
김치류	268.2	238.6	263.2	226.0	237.6
채소류	347.2	364.2	346.4	271.3	311.5
과일류	181.6	483.6	430.2	242.1	375.2
소고기류	56.8	58.6	61.9	102.3	74.4
돼지고기류	33.9	72.1	80.3	109.5	91.7
닭고기류	60.2	69.5	68.0	106.2	88.0
어류	207.3	77.6	91.0	115.0	98.7
패류/갑각류	24.1	20.9	15.7	83.8	40.0
연체류	24.2	16.6	17.0	83.8	29.4
해조류	43.4	16.3	16.7	84.5	36.7
합계	1763.4	1852.3	1852.3	1852.3	1852.3
확률	0.806	0.786	0.769	0.730	0.727

단위:g/day, N=18,714