

간이분석법을 이용한 동료평가의 공정성 향상 방안

최경호^{1,a}

^a전주대학교 기초의과학과

요약

동료평가는 일반적으로 교육 분야의 평가시스템에서 같이 학습을 하는 동료가 평가자가 되어 상대 학습자의 과정이나 결과물에 가치를 부여하는 것으로서, 많은 장점을 갖고 있음에도 불구하고 단순히 부여 받은 점수들의 평균을 계산하여 기여도를 산정한다는 단점 또한 갖고 있어 이를 개선하기 위한 다각적인 연구가 진행되어 오고 있다. 그런데 이들 선행연구에서 활용된 카이제곱 검정이나 유의확률 등의 개념은 동료평가를 활용하는 교육학이나 산업공학 분야에 종사하는 사람들에게는 쉬운 개념이 아니다. 이에 본 연구에서는 표준편차라는 간단한 통계를 이용하지만, 이들과 유사한 결과를 보이면서도 방법적인 측면에서 매우 간단한 동료평가의 가중치 부여방법을 제안하였다. 더불어 제안된 방법의 용이성과 효용성 및 활용 예에 대해서도 알아보았다. 본 연구는 간이분석법을 이용하여 이용자의 편리성을 증대시키는데 기여할 수 있다고 사료되는 바, 이러한 측면에 본 연구의 의의가 있다고 하겠다.

주요어어: 간이분석, 표준편차, 동료평가, 카이제곱검정, 유의확률.

1. 서론

불확실성을 조사하거나 실험을 통한 대량의 데이터를 얻어 분석함으로써 불확실한 가운데서 규칙성을 발견할 수 있다. 따라서 통계란 불확실성을 규명하기 위해 수집된 데이터로부터 얻어지는 정보이며, 데이터의 해석을 위한 과학적 방법을 통계적 방법이라 하고 그 방법론에 대한 지식체계를 통계학이라 한다. 이렇게 정의되는 통계학은 간단히는 평균을 얻는 방법 혹은 그림이나 도표 등을 그리는 방법에서부터 고차적인 방법인 두개골에 따른 종족분류나 암 수술환자의 수명연구 등 거의 모든 학문분야에서 지식을 생산하는 도구를 연구 개발하는 학문이라 할 수 있다. 통계학이 기여한 분야를 보면 20세기 말에 생겨나 현재 전 세계적으로 가장 연구가 활발히 진행되는 생명정보공학(bioinformatics)은 생물학과 정보 그리고 통계적 방법이 합쳐진 학문이며, 계량경제학(econometrics)은 경제학과 통계학이 합쳐져 생겨난 분야이며 그 외에도 많이 있다 (김달호 등, 2008).

한편 통계학을 이용하는 대부분의 경우에서 표본에 의하여 모집단의 특성을 추측할 때 표본이 포함하고 있는 정보를 최대한 이용하기 위하여 계산상의 어려움이 있더라도 효율이 높은 통계적 추정 및 검정방법을 사용하곤 한다. 그러나 데이터를 쉬우면서도 양적으로 많이 얻을 수 있다면, 효율이 좀 떨어지더라도 계산방법이 간단하고, 판단하기가 쉬운 방법을 사용하는 것이 바람직하다. 특히 실험이나 공장 등에서 일상적으로 대량의 데이터가 얻어지는 경우는 정밀한 수법을 사용하여 해석하기보다는, 시간이 적게 소요되는 간편한 수법으로 충분히 목적을 달성할 수 있는 경우가 많다. 이러한 간단한 분석방법을 간이분석법이라고 한다 (박성현과 박영현, 1995). 간이분석법의 한 예로 검정 시 불편분산 대신 계산이 용이한 범위를 산포의 측도로 사용하는 경우를 들 수 있다. 범위는 추정량으로서의 효율은 떨어지지만, 적절한 표본 크기만 유지되면 계산이 매우 간단하고 유용한 대안이다.

¹ (560-759) 전주시 완산구 효자동 1200, 전주대학교 기초의과학과(통계학), 교수. E-mail: ckh414@jj.ac.kr

통계학이 여러 학문분야에서 지금보다 더욱 활발하게 활용되기 위해서는, 통계학적인 측면에서의 효율성 증대도 중요하지만 사용자 입장에서의 간편성과 용이성도 고려되어야 한다. 이러한 측면에서 본 연구에서는 통계적 검정과 같은 복잡하고 어려운 통계방법 대신 표준편차와 같은 기초통계량을 활용하더라도 소기의 목적을 달성함으로써 이용자의 편리성을 증대시킬 수 있는 한 가지 간이분석방법을 소개하고자 한다.

동료평가는 일반적으로 교육 분야의 평가시스템에서 같이 학습을 하는 동료가 평가자가 되어 상대 학습자의 과정이나 결과물에 가치를 부여하는 것으로서, 전문가나 교사에 의해 평가가 전적으로 이루어지던 방식과는 다른 방식의 평가이다 (김민정, 2008). 많은 장점을 갖고 있음에도 불구하고 단순히 부여 받은 점수들의 평균을 계산하여 기여도를 산정한다는 단점 또한 갖고 있는 바, 이를 개선하기 위한 다각적인 연구가 진행되어 오고 있다. 그 한 예로 카이제곱 검정을 수행하고 이 때 도출되는 유의 확률(p -value)을 이용하여 가중치를 구해 공정성이 향상되는 동료평가 방법을 제안한 고성석과 조미연 (2010)의 연구를 들 수 있다. 그런데 고성석과 조미연 (2010)의 연구에서 활용된 카이제곱 검정이나 유의 확률 등의 개념은 동료평가를 활용하는 교육학이나 산업공학 분야에 종사하는 사람들에게는 쉬운 개념이 아니다. 이에 본 연구에서는 표준편차라는 간단한 통계를 이용하지만, 고성석과 조미연 (2010)과 유사한 결과를 보이면서도 방법적인 측면에서 매우 간단한 동료평가의 가중치 부여방법을 제안하고자 한다.

2. 동료평가 방법 고찰

오픈소스 소프트웨어를 시작으로, 집단 협업의 형태는 점점 더 다양해지고 그 적용범위는 나날이 늘어나고 있다. 기존의 전통적인 방법과는 다르게, 집단협업에서는 다양한 구성원들이 자유롭게 자신의 일을 선택할 수 있는 등 여러 가지 특성을 가지고 있기 때문에 이에 맞는 관리 방법들이 필요하다. 그 일환으로 참여자들의 참여를 유도하는 동기부여를 위해 적절한 인센티브 분배를 고려할 수 있는데, 이를 위해서는 공정하게 기여도를 측정할 수 있어야 하는데 동료평가가 한 가지 대안이 되겠다. 집단 협업에서의 동료평가란 한 집단협업에 참여한 다양한 참가자들이 각자 다른 참가자들에 대한 기여도를 직접 산정하는 방법으로서 제 3자가 아닌 참가자들이 직접 평가를 한다는 데에 의의를 두고 있다 (고성석과 조미연, 2010).

한편 이경미 (2010)에 따르면, 동료평가(peer assessment)는 참여자 중심의 평가로서 타당성을 확보하고 있으며, 감독관의 일방적인 지시보다 동료 학생들에 의한 평가에 의해 매우 많은 영향을 받게 된다. 따라서 동료평가는 감독관 1인 평가에 의한 평가의 주관성을 극복하고 평가자의 선입견을 배제할 수 있으며, 평가의 객관성과 신뢰성을 확보할 수 있는 평가방법이다. 이러한 장점에도 불구하고 평가 전에 평가대상에 대한 기준이 분명하게 설정되지 않으면 올바른 평가를 할 수 없다는 점 (Green, 1994), 동료에 대해 극단적인 평가를 하는 것을 꺼린다는 점 (Kennedy, 2005), 기여가 낮은 사람은 자신들의 낮은 기여를 위장하기 위해 비슷하게 점수를 분배하는 경향을 보인다는 점 (Herbet, 2006) 등의 단점이 있어 다양한 해결책이 모색되어 오고 있다.

지금까지 연구된 대표적인 동료평가 방법들을 정리해 보면 다음과 같다. 첫째, Kennedy (2005)가 제안한 방법으로, 이는 자신을 제외한 다른 팀원들만을 대상으로 평가하되 총점의 합이 (팀원수 - 1)100이 되도록 평가한다. 이 때 100점을 기준으로 평균보다 많은 기여를 한 팀원에게는 100점 이상을 부여하고, 기여도가 적은 팀원에게는 플러스한 만큼의 점수를 100점에서 제한한다. 개인점수는 (평균점수 × 팀점수)/100으로 계산한다 (표 1).

둘째, Parsons (2004)이 제안한 Zero-sum이 있다. 이는 자신을 포함 모든 팀원을 -12부터 +12까지의 범위로 점수를 부여하되 점수의 합이 0이 되도록 하는 방법이다. 기여도가 높은 팀원에게는 플러스

표 1: Kennedy의 동료평가 방법의 예

	A	B	C	총점
A		80	120	200
B	110		90	200
C	130	70		200
총점	240	150	210	600
평균	120	75	105	
팀 점수	20	20	20	
개인 점수	24	15	21	

표 2: Flexi-sum 방법의 예

	A	B	C	총점
A	10	8	5	30
B	9	8	6	30
C	8	9	7	30
평균	9.0	8.3	6.0	

표 3: 기본점수를 10점으로 한 Flexi-sum 방법의 예

	A	B	C	D	평균
A	12	8	11	8	9.75
B	10	14	5	11	10.00
C	8	8	14	9	9.75
D	10	10	10	12	10.50
총점	40	40	40	40	

점수를 기여도가 낮은 팀원에게는 마이너스 점수를 부여함으로써 기여도 정도를 명확하게 구별할 수 있고 평가방법이 복잡하지 않은 장점이 있다. 그러나 동료에게 마이너스 점수를 부여하는 부분에 대하여 거부감을 갖게 되어 비슷한 점수를 부여할 소지가 있다. 이에 대한 해결책은 60-sum이 대안으로 제시되고 있는데, 총점이 60점이고 자기 자신도 평가한다는 점만 제외하면 Kennedy (2005) 방법과 유사하다.

셋째, Flexi-sum 방법인데 이 방법 또한 Parsons (2004)이 제안한 방법이다. 이 방법은 임의의 기본점수를 할당한 후 기여정도에 따라 일정 점수를 더하거나 빼는 방법이다. 단 부여된 점수의 총합은 (팀원 수 × 기본점수)가 되도록 점수 부여를 한다 (표 2).

상기 방법들은 모두 개인들이 부여 받은 점수들의 평균을 이용하여 개인점수를 부여하는 방식이다. 따라서 간편하기는 하지만, 평가자들이 비슷하게 점수를 부여함으로써 공정성이 훼손되거나 기여정도에 대한 기준이 평가자마다 다르므로 인하여 일관성이 유지 않는 문제가 발생할 소지가 있다.

넷째, 카이제곱을 이용한 방법으로 고성석과 조미연 (2010)이 제안한 방법이다. 이 방법은 Flexi-sum 방법을 기반으로 하되, 점수의 타당성에 대한 평가를 기여도 산출에 반영함으로써 참여자들이 자신의 점수를 높이기 위한 조작을 할 수 없도록 하였는데, 이를 위하여 카이제곱검정을 활용하였다. 이 방법을 설명하기 위하여 기본점수를 10점으로 평가한 표 3의 예를 고려해 보자. 표 3에서 행 X는 평가대상자를, 열 Y는 평가자를 의미한다.

공정성 검사를 위하여 자기 자신이 평가한 점수를 제외한 점수의 평균을 표 4와 같이 구한다. 여기서 (X, Y)는 Y를 제외한 평가자들이 매긴 X의 점수의 평균을 의미한다. 예를 들어 (C, D)는 D를 제외한 A, B, C가 매긴 C의 점수의 평균이다.

다음으로 한 평가자가 매긴 점수와 그 평가자를 제외한 나머지 평가자들이 매긴 점수의 평균을 카

표 4: 자신의 평가를 제외한 평가점수

	A	B	C	D
A	9.00	10.34	9.33	10.33
B	10.00	8.67	11.67	9.67
C	10.33	10.33	8.33	10.00
D	10.67	10.67	10.67	10.00

표 5: 가중치를 고려한 최종 기여도

	A	B	C	D
기여도 점수	9.69	10.81	8.62	10.88

표 6: 공정한 평가 예

	A	B	C	D	표준편차
A	23	21	21	22	0.83
B	17	17	17	16	0.43
C	13	13	13	14	0.43
D	7	9	9	8	0.83
표준편차	5.83	4.47	4.47	5.00	

이제 공급점을 이용하여 검정해 보면, A, B, C, D 각각에 대한 유의수준이 0.667, 0.223, 0.046, 0.750으로 나타나 C는 공정한 평가를 하지 않은 것으로 판단된다. 이제 이러한 점을 최종 점수 산정에 반영하고자 각각의 가중치(= 유의수준/유의수준 합)를 구한다. 예컨대 이 경우 A의 가중치를 $(0.667/1.686) = 0.40$ 등과 같이 구하면, $B = 0.13$, $C = 0.03$, $D = 0.44$ 의 가중치가 구해진다. 이제 이를 이용하여 최종 기여도를 구해보면 표 5와 같다. 여기서 A의 기여도 점수 9.69는 $(12 \times 0.4 + 8 \times 0.13 + 11 \times 0.03 + 8 \times 0.44)$ 로부터 구해진다.

표 5의 기여도 점수를 이용할 경우 표 4의 경우와 일부 차이를 보이고 있다. 특히 카이제곱검정결과 공정성에 의심이 있다고 판단된 C에 대해 낮은 가중치가 부여되어 최종 기여도 점수가 13.11%나 낮아지게 되었다. 이렇듯 카이제곱검정을 이용한 고성석과 조미연 (2010)의 방법은 공정성이 증가되는 평가가 되도록 하는 장점을 갖게 된다.

3. 간이분석 방법의 제안

고성석과 조미연 (2010)의 방법은 분명 Flexi-sum 방법 등이 갖는 문제점을 보완하는 방법으로 나름대로 타당한 방법이라 할 수 있다. 더욱이 고성석과 조미연 (2010)은 동료평가가 공정하게 이루어진 경우와 공정하게 이루어지지 않은 경우의 예를 이용하여 자신들의 방법이 효용성이 있음을 주장하고 있다.

그런데 동료평가가 실제로 이루어지는 현장은 교육현장이나 산업현장 등으로, 카이제곱검정이나, 유의확률 등과 같은 고급통계가 어렵게 느껴질 수도 있다. 따라서 고성석과 조미연 (2010)과 같은 효과를 거두면서도 통계적으로 보다 간편한 방법개발이 절실히 요구된다. 이에 표준편차를 이용하여 가중치를 구하는 방법을 제안해 본다. 표준편차는 대표적인 기술통계로 누구나 손쉽게 구할 수 있으며 이해할 수 있는 통계량이다.

고성석과 조미연 (2010)이 자신의 방법의 효용성을 입증하려고 보인 두 개의 예 중 먼저 공정한 평가의 예로 제시한 표 6을 보자.

일관성 지표인 행의 표준편차는 [0.83, 0.43, 0.43, 0.83] 등으로 0에 가깝고, 공정성 지표인 열의 표

표 7: 카이제곱검정을 이용한 수정된 개인 기여도

	A	B	C	D
Flexi-sum	21.75	16.75	13.25	8.25
카이제곱검정	21.73	16.75	13.25	8.27
증감률	-0.09%	0.00%	0.00%	+0.24%

표 8: 자신의 평가를 제외한 평가점수

	A	B	C	D
A	21.33	22.00	22.00	21.67
B	16.67	16.67	16.67	17.00
C	13.33	13.33	13.33	13.00
D	8.67	8.00	8.00	8.33
표준편차	5.3453	5.8759	5.8759	5.6856
가중치	0.266	0.242	0.242	0.250

표 9: 제안된 방법을 이용한 수정된 개인 기여도

	A	B	C	D
Flexi-sum	21.75	16.75	13.25	8.25
카이제곱검정	21.73	16.75	13.25	8.27
제안된 방법	21.78	16.75	13.25	8.22
증감률	0.14%	0.00%	0.00%	-0.36%

표 10: 공정하지 않은 평가 예

	A	B	C	D	표준편차
A	23	21	21	15	3.00
B	17	17	17	15	0.87
C	13	13	13	15	0.87
D	7	9	9	15	3.00
표준편차	5.83	4.47	4.47	0.00	

준편차는 [5.83, 4.47, 4.47, 5]로 서로 비슷한 값을 가져 표 6의 평가는 공정한 평가임을 알 수 있다. 한편 카이제곱검정을 통한 유의확률을 이용하여 가중치를 구하고, 이를 토대로 수정된 개인 기여도를 산출해 보면 표 7와 같다.

표준편차를 이용하는 제안된 방법은 매우 간편하다. 먼저 표 7에 대해 카이제곱검정 방법에서와 같이 자기 자신을 제외한 평가점수를 표 8과 같이 구한다. 그리고 표준편차를 구한다. 자신의 점수를 제외하고 평가를 했을 때 표준편차가 크다는 것은 그 평가자가 산포를 작게 하는데 기여도가 크다는 뜻이다. 따라서 그 평가자의 공정성은 작다는 의미가 된다. 이에 표준편차의 역수를 구하고 전체 크기가 1이 되도록 조정하여 가중치로 하면 표 8에 나타난 바와 같다. 이제 이를 이용하여 최종 기여도를 구하고 정리하면 표 9와 같다. 미세한 차이는 있으나 복잡하게 수행된 카이제곱검정의 결과와 매우 흡사함을 알 수 있다.

다음으로 공정하게 평가가 이루어지지 않은 예로 제시된 표 10의 예를 보자. 행의 표준편차가 0에 가까운 값이 아닌바 이는 한 사람에게 부여된 점수들이 차이가 난다는 의미이고, 열의 표준편차 역시 비슷하지 않아 공정한 평가가 이루어지지 않음을 알 수 있다. 이 경우에 대해서도 앞서서와 같은 요령으로 Flexi-sum 방법, 카이제곱검정 방법, 제안된 방법에 대해 개인 기여도를 산출해 보면 표 11과 같다.

비교적 공정하게 평가하지 않은 것으로 판단되는 D의 경우 -5.2%로 -12.9%의 카이제곱검정에 비

표 11: 공정하지 않은 평가에 대한 개인 기여도 비교

	A	B	C	D
Flexi-sum	20.00	16.50	13.50	10.00
카이제곱검정	21.29	16.96	13.04	8.71
제안된 방법	20.52	16.64	13.36	9.48
증감률	+2.60%	+0.85%	-1.04%	-5.20%

표 12: 5조의 Flexi-sum 평가

	A	B	C	D	표준편차
A	12	10	10	4	1.91
B	14	12	13	18	2.63
C	6	8	8	4	1.91
D	8	10	9	4	2.63
표준편차	3.65	1.63	2.16	7.12	

표 13: Flexi-sum 점수와 제안된 방법을 이용한 최종점수

		A	B	C	D
표 12	Flexi-sum	11.50	14.25	6.50	7.75
	제안된 방법	11.54	14.26	6.46	7.74
표 13	Flexi-sum	6.75	11.00	9.75	12.50
	제안된 방법	6.97	10.76	9.24	12.92

하여 감소폭이 작고, C에 대해서는 증가 폭이 많지만 증감의 방향은 카이제곱검정의 경우와 모두 일치하고 있다. 카이제곱검정을 이용한 방법이 최상의 방법(optimal solution)이라고 단언할 수도 없지만, 실령 최상의 해결책이라 하더라도 제안된 방법은 정확성 측면에서 약간은 떨어지지만 대신 용이성을 갖는 만큼 충분히 의미 있는 대안이라 할 수 있겠다.

4. 결과의 활용 및 결론

본 연구에서 제안한 방법은 동료평가의 신뢰성 여부를 보다 쉽고 빠르게 판단하는 데에도 활용될 수 있다. 예컨대 일관성과 공정성이 확보될 때 신뢰할 수 있는 평가라 할 수 있는데, 강정은 (2010)에 제시된 다음의 두 예는 모두 일관성과 공정성이 그다지 높지 않은 경우이다.

이와 같이 동료평가의 결과가 다수 있을 때 어떤 결과가 보다 더 신뢰할 수 있는지 여부를 결정하는데 본 연구에서 제안한 방법을 활용해 보자. 이를 위해 표 12와 13에 대한 Flexi-sum 점수와 제안된 방법을 이용한 최종점수를 살펴보면 표 14와 같다. 이로부터 Flexi-sum 점수와 제안된 방법을 이용한 최종점수 간의 차이가 더 적은 5조(표 12)의 동료평가가 10조(표 13)의 동료평가 보다 더 신뢰할 수 있음을 알 수 있다.

일반적으로 동료평가 방법은 참여자 중심의 평가방법으로 객관성을 확보할 수 있는 평가방법으로 알려져 있다. 그러나 평가자들의 심리적인 문제와 주관적인 평가로 인해 공정하고 정확한 평가가 이루어지지 못하는 경우가 있어, 이에 대한 개선을 위한 연구가 진행되어 오고 있다. 그 연구의 한 예가 고성석과 조미연 (2010)의 연구이다. 그런데 이들 연구에서 제안된 방법을 활용하려면 복잡한 통계적인 검정법 등을 이용해야 하기에 현장에서 쉽게 적용하기에는 어려움이 따른다. 이에 본 연구에서는 이들 연구와 대동소이 한 결과를 도출하면서도 표준편차라는 이해하기 쉽고 구하기 쉬운 통계량을 이용하는 방법에 대해서 제안하였다. 더불어 제안된 방법의 용이성과 효용성 및 활용 예에 대해서도 알아보았다. 전술한바와 같이, 통계학이 여러 학문분야에서 지금보다 더욱 활발하게 활용되기 위해서는 통

계학적인 측면에서의 효율성 증대도 중요하지만 사용자 입장에서의 간편성과 용이성도 고려되어야 한다. 이에 본 연구는 간이분석법을 이용하여 이용자의 편리성을 증대시키는데 기여할 수 있다고 사료되는 바, 이러한 측면에 본 연구의 의의가 있다고 하겠다.

참고 문헌

- 강정은 (2010). <동료평가를 이용한 개인 기여도 평가에 관한 연구>, 건국대학교대학원 석사학위논문.
- 고성석, 조미연 (2010). <통계적 기법을 이용한 효율적 동료 평가법>, 2010년 대한산업공학회/한국경영과학회 춘계공동학술대회.
- 김달호, 이경은, 손중권, 송재기, 박병구 (2008). <생활속의 통계>, 자유아카데미.
- 김민정 (2008). 웹기반 형성적 동료평가 시스템 개발을 위한 설계기반 연구, *The Journal of Educational Information and Media*, **14**, 85-114.
- 박성현, 박영현 (1995). <통계적 품질관리>, 민영사.
- 이경미 (2010). <쓰기과업에서 동료평가가 자기평가에 미치는 영향>, 중앙대학교 교육대학원, 석사학위논문.
- Green, M. (1994). Children as evaluators: Understanding evaluation from the inside, In *Responsive Evaluation*, Brian Cambourne and Jan Tubil(eds.), Portsmouth, NH: Heinemann, 83-103.
- Herbet, N. (2006). Peer assessment: Is it fair?, In *Uniserve science assessment symposium proceeding*, Sydney.
- Kennedy, G. J. (2005). Peer-assessment in group projects: Is it worth it?, *Australian Computer Society*, **42**, 59-65.
- Parsons, D. (2004). Justice in the Classroom: Peer Assessment of Contributions in Group Projects, *Proceedings of the 17th NACCQ*, 145-151.

The Method of Improvement in Fairness on Peer Assessment - Based on Convenience Analysis

Kyoung-Ho Choi^{1,a}

^aDepartment of Basic Medical Science, Jeonju University

Abstract

Peer assessment is an educational valuation system that involves studying with a colleague and granting value to the progress made by the colleague. Although this method has many merits, there is also a drawback pertaining to calculating the mean of the scores that were granted to the levels of contribution. However, this has been improved upon by a diversified study. However, the concept of the chi-square test and p -value used in the preceding study is not easy individuals engaged in the industrial engineering field or education when using peer assessment. This study uses simple statistics like standard deviation, in addition to, investigating the availability of a suggested method as well as examples of utility and application. This study can contribute to increase the convenience of users through the use of convenience analysis and with this method.

Keywords: Convenience analysis, standard deviation, peer assessment, chi-square test, p -value.

¹ Professor, Department of Basic Medical Science(Statistics Major), Jeonju University, Jeonju 560-759, Korea.
E-mail: ckh414@jj.ac.kr