韓國數學敎育學會誌 시리즈 A <數學敎育>
2000, 11. 제 39권, 제 2호, 167-177.

J. Korea Soc. Math. Ed. Ser. A: The Mathematical Education
Nov. 2000, Nov. 39, No. 2, 167-177.

동형고사에서 표준점수차의 확률분포를 이용한 상사성의 측정과 평가치 산정에 관한 연구*

홍 석 강 (동국대학교)

Ⅰ. 서 론

동형검사는 측정하는 능력이나 행동 특성 등에서 같은 특성을 측정하기 위하여 시행하는 동질의 검사이며 일반적으로 그 검사도구들이 양호한 검사이기 위해서 고려되어야 할 중요한 요인들로 문항의 타당도, 신뢰도, 객관도, 문항 변별도 및 측정자가 세운 기준의 문항 반응분포도 등이 있다. 그 가운데 특히 신뢰도는 두 개의 동형검사를 제작하여 동일한 집단에 그 두 검사를 시행하든지 아니면 상이한 장소의 다른 수험자군에게 동형의 고사를 동시에 시행하여 그 두 검사에서 얻은 평가 결과의 상관계수를 구하는 것인데 그것은 두 검사들의 상사성의 크기를 측정하는데 있어서 매우 중요한 요인 중 하나이다. 이 때, 그 검사들의 평가치가 원점수로 표시된 경우는 그 평가치들의 점수 차가 비교적 크므로 평가 등급의 매김과 신뢰도 계수의 계산이 비교적 용이하지만 그 검사 점수를 다른 대체 점수로 변환한다든지 아니면 표준화시킨 경우 두 검사점수의 차이는 대개 작은 값으로 나타나 신뢰도 계수가 0.5이하의 값으로 계산되어 그 값들로써는 동질성의 평가와 수험자들의 평가등급에 대한 산정 및 그 검정 요인에 대한 분석이 어려운 경우가 많다. 따라서 이 연구에서는 그러한 난점을 조금이나마 해결할 수 있는 이론과 자료 처리법을 연구하고자 한다. 그 이유는 근년도에 시행한 대학수학능력시험에서 고사의 난이도를 반영한 변환표준점수제 도입의 결과 평가등급의 매김에는 약간 정확도를 기할 수 있었으나 그 점수의 특성상 상위권으로 올라갈수록 점수의 향상 폭이 적어지고 변별력도 그 만큼 떨어지는 것으로 나타나게 되며 또 매년 시행한 입시 평가에서 동질성의 신뢰도가 오히려 저하되고 있다는 지적도 있기 때문이다. 그러므로 여기서는 두 동형고사의 원 점수들을 표준화 평가치로 변환한 후 그 평가치 차의 값들로 이루어진 확률 분포와 그 성질들을 이용하여 두 동형고사의 상사성을 측정함에 있어서 신뢰도 계수를 효과적으로 구하고 수험자들의 지정된 능력에 대한 평가 등급도 공정하게 평가해 줄 수 있는 연구를 개발하고자 하는데 이것을 간단히 요약하여 기술하면 두 동형고사의 평가치의 평균치와 분산이 모두 공통된 값을 가지고 정규분포를 따른다는 가정 하에 그 평가치들을 표준화시킨 평가치들의 차의 확률 분포로써 그 검사들의 상사성에 대한 정확한 판정을 할 수 있도록 중위수로 표현된 신뢰도계수의 하한치 즉 최대상사성계수를 설정하는 방법과 그 평가치들의 차에 대한 평가등급의 확률치를 산정하는 통계적 해석법을 주 연구 내용으로 다루고저 한다.

Ⅱ. 연구의 내용

(1) 연구의 기초 이론과 그 배경

일반적으로 두 동형고사의 표준화 점수를 X, Y로 표현한 경우 X와 Y가 모두 공통의 평균치와 분산을 가지고 정규분포 $N(\mu, \sigma^2)$를 따를 때 그 두 동형검사 점수의 차 $|d|=|X-Y|$의 확률분포와 그 누적분포는 일반정규분포와는 다른 성질을 가지는 것으로 잘 알려져 있다. 이 분야의 연구에는 Daniel(1959)가 반정규분포(Half Normal Distribution) 및 그 성질을 증명한 것을 기초로 하는데 그는 표준정규분포의 원점인 평균점을 중심으로 x축의 오른편에 있는 확률밀도를 총확률밀도로

* 2000년 6월 투고, 2000년 10월 심사 완료.
* 본 연구는 2000년도 동국대학교 전문학술지 논문게재 연구비 지원으로 이루어졌음.

계산하여 그것을 반정규분포로 정의하였다. 다음 Leone, Nelson 와 Nottingham(1961)과 Elandt(1961) 등이 위의 반정규분포의 정의를 이용하여 정규분포에서 평균치와 분산의 크기가 x축에 대칭 이동되어진 점을 중심으로 정규분포의 x좌표 왼편에 있는 확률밀도들을 x좌표의 오른편에 접어 그 확률밀도의 크기를 총확률밀도로 계산하여 새로운 분포로 정의하고 그것을 접정규분포(Folded Normal Distribution)라는 이름으로 명명하였다. 그후 Plomin 과 DeFries(1980)는 정규분포의 경우 $|d|$의 평균치와 분산은 각각

$$\mu_{|x-y|}=2\sigma\sqrt{\frac{1-\rho}{\pi}}$$

$$\sigma^2_{|x-y|}=2\sigma^2(1-\rho)\left(1-\frac{2}{\pi}\right)$$

(단, ρ는 X와 Y의 상관계수.)

임을 증명하였고 McGraw 와 Wong(1994)는

$$W=\left(\frac{z}{\sqrt{2\sigma^2(1-\rho)}}\right)^2,\quad \text{단, } z=x-y.$$

는 χ^2_1 분포에 따름을 간단히 지적하고 있다. 그러므로 이 연구에서는 위의 학자들의 연구결과를 토대로 McGraw 와 Wong(1994)가 제시한 증명을 조금 더 쉬운 표현으로 증명하고 그 성질을 이용하여 표준정규분포의 경우가 아닌 T점수 등 기타 다른 표준점수의 경우 새로운 접정규분포의 확률치와 누적분포치를 계산함으로써 서론에서 요약한 평가 목적에 부합되는 평가 해석법을 고안하고자 한다.

(2) 주 연구 내용

정의 1. (표준점수) 표준점수(Standardized Score)는 점수 분포의 평균과 표준편차가 표준 값이 되도록 변환한 점수이다.

대표적인 표준점수는 <표 1>의 예와 같이 들 수 있다

<표 1> 표준점수의 예

점 수 명	평균 (μ)	표준편차 (σ)
표준화 점수	0	1
T 점수	50	10
AGCT(Army General Classification Test)	100	20
WIT(Wechsler Intelligence Test)	500	15
GRE 테스트	500	100
편차 IQ	100	15

정의 2. (접정규분포) 정규확률변수 Y의 확률밀도함수를

$$g(y)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad \cdots\cdots (1)$$

라 할 때, $X=|Y|$의 확률밀도함수는

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}+e^{-\frac{(x+\mu)^2}{2\sigma^2}}\right],$$

$$\text{단, } x\geq 0 \quad \cdots\cdots (2)$$

이며, 이 $f(x)$를 접정규 확률밀도함수(Folded Normal p.d.f.)라 한다.

여기서 접(Folded)고자 하는 기준점을 f로 표시할 때 이 분포의 평균치를 μ_f, 분산을 σ^2_f로 표기하면

$$\mu_f=\sigma\sqrt{\frac{2}{\pi}}e^{-\frac{\mu^2}{2\sigma^2}}+\mu\left[1-2F\left(-\frac{\mu}{\sigma}\right)\right]$$

$$\text{단, } F(a)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}e^{-\frac{t^2}{2}}dt. \quad \cdots\cdots (3)$$

이고

$$\sigma^2_f=\mu^2+\sigma^2-\left\{\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\sigma e^{-\frac{\mu^2}{2\sigma^2}}+\mu\left[1-2F\left(-\frac{\mu}{\sigma}\right)\right]\right\}^2 \quad \cdots\cdots (4)$$

이다.

이 식으로 μ와 σ^2의 최우추정치는

$$\sum_{i=1}^{n}\frac{x_i}{1+e^{-\frac{2\hat{\mu}x_i}{\hat{\sigma}^2}}}=\frac{\left(n\hat{\mu}+\sum_{i=1}^{n}x_i\right)}{2} \quad \cdots\cdots (5)$$

$$\sum_{i=1}^{n}\frac{x_i}{1+e^{-\frac{2\hat{\mu}x_i}{\hat{\sigma}_i^2}}}$$

$$= \frac{[\sum(x_i + \hat{\mu})^2 - \hat{\sigma}^2 n]}{4\hat{\mu}} \cdots\cdots (6)$$

이므로 식 (5)와 (6)을 이용하여

$$\frac{\left(\frac{\mu_f}{\sigma_f}\right)^2}{1+\frac{\mu_f}{\sigma_f}} = G(\theta)$$

$$\sigma^2 = \frac{(\mu_f^2 + \sigma_f^2)}{1+\theta^2} G(\theta)$$

$$= \frac{\sqrt{\frac{2}{\pi}} e^{-\frac{\theta^2}{2}} + \theta[1-2F(-\theta)]}{1+\theta^2}$$

$$\text{단, } \theta = \frac{\mu}{\sigma} \cdots\cdots (7)$$

를 유도할 수 있고, 이 식으로써 Leone, Nelson 과 Nottingham(1961)[1]은 그들의 논문에서 도표Ⅱ와 표Ⅱ의 수표제작과정을 제시하고 있다.

정리 1. [Plomin 과 DeFries(1980)의 정리]

X와 Y가 공통의 평균치 μ와 분산 σ^2을 갖는 정규확률변수라 하고 그 상관계수가 ρ라 하면 $Z=|X-Y|$의 평균치와 분산은

$$\mu_z = 2\sigma\sqrt{\frac{1-\rho}{\pi}} \cdots\cdots (8)$$

$$\sigma_z^2 = 2\sigma^2(1-\rho)\left(1-\frac{2}{\pi}\right) \cdots\cdots (9)$$

이다.

정의 3. [McGraw 와 Wong(1994)의 최대상사성계수 Med (| d |)]

$$\text{Med}(|d|) = me(|d|)\,\sigma\sqrt{1-r} \cdots\cdots (10)$$

단, $me(|d|)$는 표준정규분포의 경우 $|d|$ 의 중위수.

한편 위의 정리 1의 전제 하에 McGraw와 Wong(1994)가 지적한 증명을 조금 더 쉬운 표현으로 정규분포인 z분포에서

$$W = \frac{Z^2}{2\sigma^2(1-\rho)}$$

로 변환하여 χ_1^2분포에 따름을 증명하고 그 경우 평균치와 분산은 역시 식 (8)과 (9)와 같음을 보일 수 있다.

정리 2. $Z=|X-Y|$일 때,

$W = \frac{Z^2}{2\sigma^2(1-\rho)}$는 χ_1^2분포에 따르며 Z의 평균치와 분산은 식 (8)과 (9)와 같다.

(증명)

X와 Y가 모두 공통인 평균치 μ와 분산 σ^2을 가지므로 $X-Y$는

평균치 $\mu_{(x-y)} = \mu_x - \mu_y = 0$,

$$\text{분 산 } \sigma^2_{(x-y)} = \sigma^2_{(x)} + \sigma^2_{(y)} - 2\rho\,\sigma(x)\sigma(y)$$

$$= 2\sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1-\rho)$$

를 가진다. 여기서

$$W = \frac{Z^2}{2\sigma^2(1-\rho)} = \left[\frac{(X-Y)}{\sqrt{2\sigma^2(1-\rho)}}\right]^2$$

이므로 W은 χ_1^2분포에 따름을 알 수 있고 그 평균치와 분산은 다음과 같이 구해진다.

$$E(\sqrt{w}) = \int_0^\infty w^{\frac{1}{2}} \frac{1}{\Gamma_{(\frac{1}{2})} 2^{\frac{1}{2}}} w^{\frac{1}{2}-1} e^{-\frac{w}{2}} dw$$

$$= \int_0^\infty \frac{1}{\Gamma_{(\frac{1}{2})} 2^{\frac{1}{2}}} e^{-\frac{w}{2}} dw$$

$$= \frac{1}{\Gamma_{(\frac{1}{2})} 2^{\frac{1}{2}}} \int_0^\infty e^{-\frac{w}{2}} dw$$

$$= \frac{1}{\Gamma_{(\frac{1}{2})} 2^{\frac{1}{2}}} 2 = \frac{\sqrt{2}}{\Gamma_{(\frac{1}{2})}} = \frac{\sqrt{2}}{\sqrt{\pi}}$$

이고, 따라서

$$\mu_z = E(z) = \sqrt{2\sigma^2(1-\rho)}\,E(\sqrt{w}) = 2\sigma\sqrt{\frac{1-\rho}{\pi}}$$

1) 수표제작의 결과는 그들의 연구내용에서 도표Ⅱ와 표Ⅱ를 참고할 것.

이다. 또, $E(z^2)=2\sigma^2(1-\rho)E(w)$이므로

$$E(w)=\int_0^\infty w\frac{1}{\Gamma_{(\frac{1}{2})}2^{\frac{1}{2}}}w^{\frac{1}{2}-1}e^{-\frac{w}{2}}dw$$

$$=\int_0^\infty\frac{1}{\Gamma_{(\frac{1}{2})}2^{\frac{3}{2}}}w^{\frac{3}{2}-1}e^{-\frac{w}{2}}dw$$

$$=1$$

이고

$$E(z^2)=2\sigma^2(1-\rho)$$

이다. 따라서

$$\sigma_z^2=E(z^2)-\mu_z^2=2\sigma^2(1-\rho)-4\sigma^2\left(\frac{1-\rho}{\pi}\right)$$

$$=2\sigma^2(1-\rho)\left(1-\frac{2}{\pi}\right)$$

이고 그 결과는 식 (8)과 (9)와 동일하다.

다음의 <표 2>는 McGraw 와 Wong(1994)가 제시한 결과로 표준정규분포의 경우 상관계수 ρ와 표준편차 σ로써 $|d|$의 수치와 관련된 확률하한치 및 확률상한치를 계산한 것이다.[2)]

지금 정리 2를 이용하여 <표 2>와 같은 확률하한치와 확률상한치를 계산하는 경우 그 하한은 z축에서

$$z^*=\frac{|d|}{\sqrt{2\sigma^2(1-\rho)}}$$

인 점을 계산한 후 $z=0$에서 $z=z^*$사이의 정규확률 밀도의 크기를 계산하든지 아니면 $\chi_1^2=\frac{|d|^2}{2\sigma^2(1-\rho)}$를 이용하여 χ_1^2 분포의 밀도의 크기로 그것을 측도 할 수 있다. 이때 그 상한 확률치의 크기는 그 하한 확률치의 여 확률밀도의 크기인데 그것은 z^*이상의 확률밀도 크기로 계산할 수 있으며 그 경우 두 개의 확률치 모두 χ_1^2의 확률밀도 크기가 z^* 확률밀도 크기의 2배로 도시된다. 또, Jensen(1970)와 McCall(1979)는 (10)식의 상사성의 측도에서 IQ 평가치와 같이 평가치의 차가 매우 작은 경우 중위수를 이용한 상관계수를 이용할 것을 제안

<표 2> 표준정규분포의 경우 $|d|$의 확률 하한치 (Lower Tail Probability) 및 확률 상한치(Upper Tail Probability)

$\sigma\sqrt{(1-\rho)}$단위상에서의 $\lvert d\rvert$값	확률하한치	확률상한치
0.008862	0.005	0.995
0.017725	0.010	0.990
0.044319	0.025	0.975
0.088681	0.050	0.950
0.177712	0.100	0.900
0.450624	0.250	0.750
0.953873 중위수	0.500	0.500
1.128379 평균치	0.575	0.425
1.626837	0.750	0.250
2.326173	0.900	0.100
2.771808	0.950	0.050
3.169823	0.975	0.025
3.642774	0.990	0.010
3.969746	0.995	0.005
4.653601	0.999	0.001

하고 있으며 이 논문의 내용처럼 T 점수의 경우 각 T 점수들의 표준정규변량 Z_x와 Z_y가 모두 완전한 표준정규분포인 Z분포에 수렴한다면 Ⅲ절 <표 5>의 분포를 이용하는 것이 바람직하나 현실적으로 그런 자료를 얻기가 어려울 때는 직접 표준점수 분포의 확률밀도를 계산하여 중위수를 구하고 그 표준점수의 상관계수를 이용해도 무방함을 지적하고 있다. 이와 같이 앞의 여러 학자들이 제시한 바의 접정규분포의 성질을 이용한 $\mu_{|d|}$와 $\sigma_{|d|}$의 지정법과 그 누적확률 분포의 계산과정 및 수표 제작법들은 그 설명 내용이 너무 어렵고 그 시행 방법도 평가 현장에 직접 적용하기에는 적당하지 않은 면이 많으므로 그 자료처리법을 좀 더 쉽게 기술하고 확률분포치의 계산 및 수표작성법에 관한 구체적인 예를 들어 보임으로써 수학교육평가 현장에서 누구나 쉽게 이 연구결과들을 이용할 수 있도록 하고자 한다.

2) McGraw 와 Wong(1994) P105, Table 1 에서 인용한 것 임

Ⅲ. 계산 예 및 검토

아래 자료는 어느 대학 입시 평가 연구소에서 시행한 고등학교 3학년 월말고사의 수학 성적들인데 이 때 두 반의 학생들을 지적 능력과 정의적 능력이 모두 동일한 수준인 통제 집단으로 두었다. 이 경우 그 고사 점수들로써 두 학급의 상사성의 크기를 측정하기 위하여 최대 상사성 계수 및 확률의 하한치와 그 상한치를 산정하는 과정과 통계적 해석법을 다음과 같은 예로 제시하였다.

여기서 같은 번호를 갖는 학생들을 한 조로 하여 두 반의 상관계수를 구한 것은 <표 4>(나)인데 이 값들은 모두 작은 것으로서 상관관계가 거의 없는 것으로 나타나고 있는 바 그 값들로써 신뢰도 계수의 판정이 매우 어려우므로 다음과 같이 신뢰도 계수 검정을 위한 후속적인 조치가 필요하다. 지금 A반과 B반의 두 학급을 한 조로 하여 T점수를 구하고 T_1과 T_2를 자료로 하여 $|d|=|T_1-T_2|$ 를 계산할 경우 그 표본의 총 크기는 모두 $S_1\times S_2=50\times50=2500$인데 그 $|d|$의 값으로써 정리 1의 (8)과 (9)식을 이용하여 평균치와 분산을 계산한 결과

$$\mu_{|d|}=2\sigma\sqrt{\frac{1-r}{\pi}} \quad \text{단, } \sigma=10$$

$$=2\times10\sqrt{\frac{1+0.05025}{\pi}}=11.5638.$$

$$\sigma^2_{|d|}=2\sigma^2(1-r)\left(1-\frac{2}{\pi}\right)$$

$$=2\times10^2\times(1+0.05025)\times\left(1-\frac{2}{\pi}\right)$$

$$=76.828$$

즉, $\sigma_{|d|}=8.765158$ 를 얻었다.

여기서 Leone, Nelson 과 Nottingham(1961)에 수록된 수표 Table 1 을 참고하였으나 $\frac{\mu_{|d|}}{\sigma_{|d|}}=1.31929$

<표 3> 원점수 자료, X(A반), Y(B반)

X(A반)

번호	1	2	3	4	5	6	7	8	9	10
	6	19	15	17	14	13	13	7	14	21
번호	11	12	13	14	15	16	17	18	19	20
	19	16	15	20	14	6	15	12	17	14
번호	21	22	23	24	25	26	27	28	29	30
	19	16	8	12	18	16	12	8	13	10
번호	31	32	33	34	35	36	37	38	39	40
	18	12	12	14	8	12	15	16	24	12
번호	41	42	43	44	45	46	47	48	49	50
	18	16	16	12	16	12	10	12	14	14

Y(B반)

번호	1	2	3	4	5	6	7	8	9	10
	28	13	14	14	28	21	12	9	10	20
번호	11	12	13	14	15	16	17	18	19	20
	30	9	23	22	19	22	8	12	17	14
번호	21	22	23	24	25	26	27	28	29	30
	14	18	19	15	14	17	22	8	18	22
번호	31	32	33	34	35	36	37	38	39	40
	21	21	16	11	14	21	31	28	10	22
번호	41	42	43	44	45	46	47	48	49	50
	9	17	11	17	17	18	10	20	13	18

<표 4> A반과 B반의 T점수와 각학급조들의 중위수 및 상관계수

(가) A반과 B반의 T점수

T_1(A반)

번호	1	2	3	4	5	6	7	8	9	10
	29.9953	62.9581	52.5080	57.7331	49.8955	47.2830	47.2830	31.6078	49.8955	8.1832
번호	11	12	13	14	15	16	17	18	19	20
	62.9581	55.1205	52.5080	65.5706	49.8955	28.9953	52.5080	44.6705	57.7331	49.8955
번호	21	22	23	24	25	26	27	28	29	30
	62.9581	55.1205	34.2204	44.6705	60.3456	55.1205	44.6705	34.2204	47.2830	39.4454
번호	31	32	33	34	35	36	37	38	39	40
	60.3456	44.6705	44.6705	49.8955	34.2204	44.6705	52.5080	55.1205	76.0207	44.6705
번호	41	42	43	44	45	46	47	48	49	50
	60.3456	55.1205	55.1205	44.6705	55.1205	44.6705	39.4454	44.6705	49.8955	49.8955

T_2(B반)

번호	1	2	3	4	5	6	7	8	9	10
	68.4221	42.5484	44.2733	44.2733	68.4221	56.3477	40.8235	35.6487	52.8979	54.6228
번호	11	12	13	14	15	16	17	18	19	20
	71.8719	35.6487	59.7975	58.0726	52.8979	58.0726	33.9238	40.8235	49.4480	44.2733
번호	21	22	23	24	25	26	27	28	29	30
	44.2733	51.1729	52.8979	45.9982	44.2733	49.4480	58.0726	33.9238	51.1729	58.0726
번호	31	32	33	34	35	36	37	38	39	40
	56.3477	56.3477	47.7231	39.0986	44.2733	56.3477	73.5968	68.4221	37.3736	58.0726
번호	41	42	43	44	45	46	47	48	49	50
	35.6487	49.4480	39.0986	49.4480	49.4480	51.1729	37.3736	54.6228	42.5484	51.1729

단, $T_1 = 50 + 10Z_x$, $T_2 = 50 + 10Z_y$

(나) 각 학급조들의 중위수 및 상관계수

한조의 반 / 중위수및상관계수(r)	A 와 B	C 와 D	E 와 F	G 와 H	H 와 I
중위 수(me)	11.650	13.921	12.862	13.014	14.038
상관계수(r)	-0.05025	0.05033	0.02193	-0.03151	0.01024

인 경우는 그들의 표기인 접정규분포에서의 값 $\frac{\mu_f}{\sigma_f}$의 수치는 최하의 값으로써 1.3236으로만 계산되어져 있으며 그 이하의 수치는 수록되지 않았으므로 그들의 연구결과들을 이용할 수 없었다. 그리고 Table 2 에서는 $\frac{\mu_f}{\sigma_f} = 1.7$인 경우 그 접정규분포의 확률하한치와 확률상한치의 수표만 작성, 제시되고 있으므로 실제 학습평가 현장에서는 이 수표가 거의 이용되어 질 수 없으며 또 그 수표 제작과정이 매우 어려워 독자들이 쉽게 접근할 수 없는 난점이 있었다. 그러므로 여기서는 정의 2의 접정규 확률 함수 $f(|d|)$

$$= \frac{1}{\sigma_{|d|}\sqrt{2\pi}}\left[e^{-\frac{(|d|-\mu)^2}{2\sigma_{|d|}^2}} + e^{-\frac{(|d|+\mu)^2}{2\sigma_{|d|}^2}}\right]$$

를 이용하여 $|d| = t_f$로 두고 확률하한치로써 누적분포함수 $F(t_f) = \int_0^{f} f(|d|)\, d(|d|)$를 직접 계산하였는데, 이때 t_f는 $|d|$의 크기를 0.0, 0.1, 0.2, …, 57 로 x좌표축에 중분으로 두고 $f(|d|) \times t_f$로써 리만적분 계산하는 방법으로 구하였다.

<표 5> T 점수로써 구한 $|d|$의 확률하한치와 확률상한치

$\sqrt{(1-\rho)}$ 단위상의 $\mid d \mid$의 값	확률하한치	확률상한치	$\sqrt{(1-\rho)}$ 단위상의 $\mid d \mid$의 값	확률하한치	확률상한치
0.08862	0.005	0.995	16.26837	0.750	0.250
0.17725	0.010	0.990	23.26173	0.900	0.100
0.44319	0.025	0.975	27.71808	0.950	0.050
0.88681	0.050	0.950	31.69823	0.975	0.025
1.77712	0.100	0.900	36.42774	0.990	0.010
4.50624	0.250	0.750	39.69746	0.995	0.005
9.53873(중위수)	0.500	0.500	46.53601	0.999	0.001
11.28379(평균치)	0.575	0.425			

<표 6> A반과 B반의 한 조에서 계산된 $|d|$의 $F(t_f)$.

단, $F(t_f)$는 확률의 하한치 임.

$\mid d \mid$	$P(X \le \mid d \mid)$	$\mid d \mid$	$P(X \le \mid d \mid)$	$\mid d \mid$	$P(X \le \mid d \mid)$	$\mid d \mid$	$P(X \le \mid d \mid)$	$\mid d \mid$	$P(X \le \mid d \mid)$
0	0.0000	4.4	0.1723	8.8	0.3660	13.2	0.5720	17.6	0.7548
0.1	0.0038	4.6	0.1805	8.9	0.3706	13.3	0.5765	17.7	0.7584
0.2	0.0076	4.6	0.1805	9	0.3753	13.4	0.5811	17.8	0.7619
0.3	0.0114	4.7	0.1847	9.1	0.3800	13.5	0.5856	17.9	0.7655
0.4	0.0152	4.8	0.1889	9.2	0.3846	13.6	0.5902	17.9	0.7655
0.5	0.0190	4.9	0.1930	9.3	0.3893	13.7	0.5947	18	0.7690
0.6	0.0228	5	0.1972	9.4	0.3940	13.8	0.5992	18.1	0.7725
0.7	0.0266	5.1	0.2014	9.5	0.3987	13.9	0.6036	18.2	0.7759
0.8	0.0304	5.2	0.2057	9.6	0.4034	14	0.6081	18.3	0.7793
0.9	0.0343	5.3	0.2099	9.7	0.4081	14.1	0.6125	18.4	0.7827
1	0.0381	5.4	0.2141	9.8	0.4128	14.2	0.6170	18.5	0.7861
1.1	0.0419	5.5	0.2184	9.9	0.4175	14.3	0.6214	18.6	0.7894
1.2	0.0457	5.6	0.2227	10	0.4222	14.4	0.6258	18.7	0.7927
1.3	0.0496	5.7	0.2270	10.1	0.4269	14.5	0.6302	18.8	0.7960
1.4	0.0534	5.8	0.2313	10.2	0.4316	14.6	0.6345	18.9	0.7992
1.5	0.0572	5.9	0.2356	10.3	0.4363	14.7	0.6389	19	0.8024
1.6	0.0611	6	0.2399	10.4	0.4410	14.8	0.6432	19.1	0.8056
1.7	0.0649	6.1	0.2443	10.5	0.4458	14.9	0.6475	19.2	0.8087
1.8	0.0688	6.2	0.2486	10.6	0.4505	15	0.6518	19.3	0.8118
1.9	0.0727	6.3	0.2530	10.7	0.4552	15.1	0.6560	19.4	0.8149
2	0.0765	6.4	0.2574	10.8	0.4599	15.2	0.6603	19.5	0.8180
2.1	0.0804	6.5	0.2617	10.9	0.4646	15.3	0.6645	19.6	0.8210
2.2	0.0843	6.6	0.2661	11	0.4694	15.4	0.6687	19.7	0.8240
2.3	0.0882	6.7	0.2706	11.1	0.4741	15.5	0.6729	19.8	0.8269
2.4	0.0921	6.8	0.2750	11.2	0.4788	15.6	0.6770	19.9	0.8298
2.5	0.0960	6.9	0.2794	11.3	0.4835	15.7	0.6811	20	0.8327
2.6	0.0999	7	0.2839	11.4	0.4882	15.8	0.6852	20.1	0.8356
2.7	0.1039	7.1	0.2884	11.5	0.4929	15.9	0.6893	20.2	0.8384
2.8	0.1078	7.2	0.2928	11.6	0.4976	16	0.6934	20.3	0.8412
2.9	0.1118	7.3	0.2973	11.7	0.5023	16.1	0.6974	20.4	0.8439
3	0.1157	7.4	0.3018	11.8	0.5070	16.2	0.7014	20.5	0.8467
3.1	0.1197	7.5	0.3064	11.9	0.5117	16.3	0.7054	20.6	0.8494
3.2	0.1237	7.6	0.3109	12	0.5164	16.4	0.7094	20.7	0.8520
3.3	0.1277	7.7	0.3154	12.1	0.5211	16.5	0.7133	20.8	0.8547
3.4	0.1317	7.8	0.3200	12.2	0.5258	16.6	0.7172	20.9	0.8573
3.5	0.1357	7.9	0.3245	12.3	0.5304	16.7	0.7211	21	0.8598
3.6	0.1397	8	0.3291	12.4	0.5351	16.8	0.7249	21.1	0.8624
3.7	0.1437	8.1	0.3337	12.5	0.5397	16.9	0.7288	21.2	0.8649
3.8	0.1478	8.2	0.3383	12.6	0.5444	17	0.7326	21.3	0.8673
3.9	0.1518	8.3	0.3429	12.7	0.5490	17.1	0.7363	21.4	0.8698
4	0.1559	8.4	0.3475	12.8	0.5536	17.2	0.7401	21.5	0.8722
4.1	0.1600	8.5	0.3521	12.9	0.5582	17.3	0.7438	21.6	0.8746
4.2	0.1641	8.6	0.3567	13	0.5628	17.4	0.7475	21.7	0.8769
4.3	0.1682	8.7	0.3613	13.1	0.5674	17.5	0.7511	21.8	0.8792

$\|d\|$	$P(X \le \|d\|)$
21.9	0.8815
22	0.8838
22.1	0.8860
22.2	0.8882
22.3	0.8904
22.4	0.8925
22.5	0.8946
22.6	0.8967
22.7	0.8987
22.8	0.9007
22.9	0.9027
23	0.9047
23.1	0.9066
23.2	0.9085
23.3	0.9104
23.4	0.9122
23.5	0.9140
23.6	0.9158
23.7	0.9175
23.8	0.9193
23.9	0.9210
24	0.9226
24.1	0.9243
24.2	0.9259
24.3	0.9275
24.4	0.9291
24.5	0.9306
24.6	0.9321
24.7	0.9336
24.8	0.9351
24.9	0.9365
25	0.9379
25.1	0.9393
25.2	0.9407
25.3	0.9420
25.4	0.9433
25.5	0.9446
25.6	0.9459
25.7	0.9471
25.8	0.9483
25.9	0.9495
26	0.9507
26.1	0.9519
26.2	0.9530
26.3	0.9541
26.4	0.9552
26.5	0.9563
26.6	0.9573
26.7	0.9584
26.8	0.9594
26.9	0.9604
27	0.9613
27.1	0.9623
27.2	0.9632
27.3	0.9641
27.4	0.9650
27.5	0.9659
27.6	0.9667
27.7	0.9676
27.8	0.9684
27.9	0.9692
28	0.9700
28.1	0.9708
28.2	0.9715
28.3	0.9722
28.4	0.9730
28.5	0.9737
28.6	0.9744
28.7	0.9750
28.8	0.9757
28.9	0.9763
29	0.9770

$\|d\|$	$P(X \le \|d\|)$
29.1	0.9776
29.2	0.9782
29.3	0.9788
29.4	0.9794
29.5	0.9799
29.6	0.9805
29.7	0.9810
29.8	0.9815
29.9	0.9820
30	0.9825
30.1	0.9830
30.2	0.9835
30.3	0.9840
30.4	0.9844
30.5	0.9849
30.6	0.9853
30.7	0.9857
30.8	0.9861
30.9	0.9865
31	0.9869
31.1	0.9873
31.2	0.9877
31.3	0.9880
31.4	0.9884
31.5	0.9887
31.6	0.9890
31.7	0.9894
31.8	0.9897
31.9	0.9900
32	0.9903
32.1	0.9906
32.2	0.9909
32.3	0.9911
32.4	0.9914
32.5	0.9917
32.6	0.9919
32.7	0.9922
32.8	0.9924
32.9	0.9927
33	0.9929
33.1	0.9931
33.2	0.9933
33.3	0.9935
33.4	0.9937
33.5	0.9939
33.6	0.9941
33.7	0.9943
33.8	0.9945
33.9	0.9947
34	0.9948
34.1	0.9950
34.2	0.9952
34.3	0.9953
34.4	0.9955
34.5	0.9956
34.6	0.9958
34.7	0.9959
34.8	0.9960
34.9	0.9962
35	0.9963
35.1	0.9964
35.2	0.9965
35.3	0.9967
35.4	0.9968
35.5	0.9969
35.6	0.9970
35.7	0.9971
35.8	0.9972
35.9	0.9973
36	0.9974
36.1	0.9975
36.2	0.9975

$\|d\|$	$P(X \le \|d\|)$
36.3	0.9976
36.4	0.9977
36.5	0.9978
36.6	0.9979
36.7	0.9979
36.8	0.9980
36.9	0.9981
37	0.9982
37.1	0.9982
37.2	0.9983
37.3	0.9983
37.4	0.9984
37.5	0.9985
37.6	0.9985
37.7	0.9986
37.8	0.9986
37.9	0.9987
38	0.9987
38.1	0.9988
38.2	0.9988
38.3	0.9988
38.4	0.9989
38.5	0.9989
38.6	0.9990
38.7	0.9990
38.8	0.9990
38.9	0.9991
39	0.9991
39.1	0.9991
39.2	0.9992
39.3	0.9992
39.4	0.9992
39.5	0.9993
39.6	0.9993
39.7	0.9993
39.8	0.9993
39.9	0.9994
40	0.9994
40.1	0.9994
40.2	0.9994
40.3	0.9994
40.4	0.9995
40.5	0.9995
40.6	0.9995
40.7	0.9995
40.8	0.9995
40.9	0.9996
41	0.9996
41.1	0.9996
41.2	0.9996
41.3	0.9996
41.4	0.9996
41.5	0.9996
41.6	0.9997
41.7	0.9997
41.8	0.9997
41.9	0.9997
42	0.9997
42.1	0.9997
42.2	0.9997
42.3	0.9997
42.4	0.9997
42.5	0.9998
42.6	0.9998
42.7	0.9998
42.8	0.9998
42.9	0.9998
43	0.9998
43.1	0.9998
43.2	0.9998
43.3	0.9998
43.4	0.9998

$\|d\|$	$P(X \le \|d\|)$
43.5	0.9998
43.6	0.9998
43.7	0.9998
43.8	0.9998
43.9	0.9998
44	0.9998
44.1	0.9999
44.2	0.9999
44.3	0.9999
44.4	0.9999
44.5	0.9999
44.6	0.9999
44.7	0.9999
44.8	0.9999
44.9	0.9999
45	0.9999
45.1	0.9999
45.2	0.9999
45.3	0.9999
45.4	0.9999
45.5	0.9999
45.6	0.9999
45.7	0.9999
45.8	0.9999
45.9	0.9999
46	0.9999
46.1	0.9999
46.2	0.9999
46.3	0.9999
46.4	0.9999
46.5	0.9999
46.6	0.9999
46.7	0.9999
46.8	0.9999
46.9	0.9999
47	0.9999
47.1	0.9999
47.2	0.9999
47.3	0.9999
47.4	0.9999
47.5	0.9999
47.6	0.9999
47.7	0.9999
47.8	0.9999
47.9	0.9999
48	0.9999
48.1	0.9999
48.2	0.9999
48.3	0.9999
48.4	0.9999
48.5	0.9999
48.6	0.9999
48.7	0.9999
48.8	0.9999
48.9	0.9999
49	0.9999
49.1	0.9999
49.2	0.9999
49.3	0.9999
49.4	0.9999
49.5	0.9999
49.6	0.9999
49.7	0.9999
49.8	0.9999
49.9	0.9999
50	0.9999
50.1	0.9999
50.2	0.9999
50.3	0.9999
50.4	0.9999
50.5	0.9999
50.6	0.9999

$\|d\|$	$P(X \le \|d\|)$
50.7	0.9999
50.8	0.9999
50.9	0.9999
51	0.9999
51.1	0.9999
51.2	0.9999
51.3	0.9999
51.4	0.9999
51.5	0.9999
51.6	0.9999
51.7	0.9999
51.8	0.9999
51.9	0.9999
52	0.9999
52.1	0.9999
52.2	0.9999
52.3	0.9999
52.4	0.9999
52.5	0.9999
52.6	0.9999
52.7	0.9999
52.8	0.9999
52.9	0.9999
53	0.9999
53.1	0.9999
53.2	0.9999
53.3	0.9999
53.4	0.9999
53.5	0.9999
53.6	0.9999
53.7	0.9999
53.8	0.9999
53.9	0.9999
54	0.9999
54.1	0.9999
54.2	0.9999
54.3	0.9999
54.4	0.9999
54.5	0.9999
54.6	0.9999
54.7	0.9999
54.8	0.9999
54.9	0.9999
55	0.9999
55.1	0.9999
55.2	0.9999
55.3	0.9999
55.4	0.9999
55.5	0.9999
55.6	0.9999
55.7	0.9999
55.8	0.9999
55.9	0.9999
56	0.9999
56.1	0.9999
56.2	0.9999
56.3	0.9999
56.4	0.9999
56.5	0.9999
56.6	0.9999
56.7	0.9999
56.8	0.9999
56.9	0.9999
57	1.0000

<표 7> 표6에서 중요한 $|d|$의 값과 확률 하한치 및 확률상한치를 수록한 표

$\|d\|$의 값	확률하한치	확률상한치
0	0.00003	0.99997
3	0.11576	0.88424
6	0.23996	0.76004
9	0.37534	0.62466
11.65(중위수)	0.50000	0.50000
12	0.516458	0.48354
12.32(평균치)	0.53047	0.46953
15	0.65181	0.34819
18	0.76902	0.23098
21	0.85987	0.14013
24	0.92269	0.07731
27	0.96416	0.03584
30	0.98258	0.01742
33	0.99293	0.00707
36	0.99742	0.00258
39	0.99915	0.00085
42	0.99975	0.00025
45	0.99993	0.00007
48	0.99998	0.00002
52	0.99999	0.00001
57	1.00000	0.00000

이와 같이 위에서 기술한 쉬운 기법을 통하여 얻은 <표 6>의 수치로써 중요한 $|d|$값의 확률하한치와 확률상한치를 요약하여 <표 7>을 얻었으며 <표 4>의 (나)에 수록된 자료로써 상관계수와 중위수를 정의 3에서 기술한 최대상사성계수 $\text{Med}(|d|) = me(|d|)\,\sigma\sqrt{1-r}$에 대입하여 두 고사의 상사성 계수의 크기를 각각 계산한 결과 그 수치들은 <표 8>과 같았다.

<표 8> 각 학급조의 상사성계수의 크기

한조의 반	상사성의 크기
A와 B	119.3011
C와 D	135.6883
E와 F	127.3272
G와 H	132.0776
H와 I	139.3360

그런데 여기서 지적하고자 하는 것은 A와 B반의 한 조의 T점수의 차이인 $\mu_{|d|} = 11.563$이었으나 실제로 $f(|d|)$의 확률분포로써 계산한 평균치는 <표 7>에서 표기한 바와 같이 12.32로 나타나고 있는 바 그 차이는 0.757인데 그것은 실제 자료인 $|d|$ 값들로 직접 구한 평균치와 리만적분을 이용한 확률밀도 크기인 $f(|d|)$의 수치적분치의 계산결과가 조금 다르게 나타나고 있음을 보여주고 있다. 이 논문에서는 기타 다른 한 조의 학급의 T점수로써 구한 평균치와 중위수도 모두 확률분포에 의하여 구한 것임을 부기한다.

이제 그 결과로써 상사성 크기 계수치의 최소하한은 119.3011로써 A반과 B반의 상사성 계수치가 최소하한치 즉, 최대상사성계수로 인정되므로 A반과 B반 두 학급의 경우가 상사성이 가장 큰 것으로 판정 되어진다. 다음에는 계산결과의 검토로써 위에서 얻은 평가 결과의 계산치들을 평가자들이 세운 평가기준과 목적에 따라 어떻게 기술할 수 있는지를 살펴 보기로 하자. 지금 그 한 예로써 <표 4>의 자료들이 T점수들을 각각 동일한 학급에 2회에 걸친 동형고사의 평가치로 가정할 경우 그 $|d|$의 점수에 대한 평가에 대한 해석을 다음과 같이 기술할 수 있을 것인데,

첫째, $|d|$가 거의 0인 경우, 즉 2회의 동형고사에서 수험자들의 점수차가 거의 차이가 없이 동일한 점수를 획득한 경우 평가자는 두 동형고사에서 상사성의 크기에 대한 유의성을 검정할 수 있는가? 아니면 점수차가 많은 경우 두 고사의 상사성에 차이가 있는지의 유의성을 검정할 수 있는가?

둘째, 동일한 교과 범위 및 수준에서 문제를 출제하였음에도 불구하고 위의 경우에서처럼 상사성의 차이 검정에 대한 유의성이 있다면 왜 수험생들의 학력에 차이가 나게 되었는가? 그 요인은 환경적 요인, 시간의 경과 및 반복적인 고사의 시행으로 인한 기억력의 효과인가? 아니면 평가자의 불일치한 평가 결과들 때문인가?

셋째, 표준화 점수 차의 등급산정을 위한 확률하한치나 확률상한치는 어떻게 해석될 수 있는가? 즉, 이 예에서 $|d|=0$에서 $|d|=57$의 확률하한치로써 평가치 차이가 없는 학생이 평가치의 차이가 큰 학생보다 우수하다고 판정할 수 있는가? 다시 말해서 이 고사들이 만일 바람직한 동형고사라면 학생들이 2회에 걸쳐 동일한

성적을 얻는 것이 오히려 바람직한 것으로 인정될 것이며 아니면 그럼에도 불구하고 현저한 성적의 차이가 있는 것으로 인정된다면 앞에서 논한 여러 가지 요인의 효과를 재검정하고 $|d|$의 등급인 확률의 하한치를 평가자의 평가기준대로 기술해야 할 것이다.

결론적으로 위의 예와 검토에서 논한 바처럼 우리 평가자는 그 평가목적과 기준에 따라 여러 가지 평가요인을 유의성 검정의 기초로 두었을 때 두 고사의 상사성에 대한 측정과 그 해석이 각각 달라질 수 있으나 $|d|$의 값에 대한 상사성의 진단으로써 최대상사성계수를 산정할 수 있고 그 확률치들을 직접적으로 계산하는 방법을 효과적으로 이용함으로써 평가 현장에서 동형고사의 상사성 측정에 유효한 평가해석을 시행할 수 있을 것이다.

Ⅳ. 결 론

본고에서는 수학교과에서 2회 시행한 동형고사의 평가결과로써 표준화 점수차의 확률분포를 이용하여 그 고사들의 상사성의 측정 및 진단에 요하는 최대상사성계수의 산정과 평가치의 등급 산정을 위한 확률하한치 및 확률상한치를 설정하는 통계적 해석법을 논하였는데 이 논문의 주요 연구 결과와 기대효과 및 활용방안을 요약하면 다음과 같다.

1. 총점을 위주로 하는 원점수의 평가와 달리 평가해석에서 일반적으로 널리 이용되고 있는 표준화 점수들의 경우 두 동형고사의 상사성 측정에서 신뢰도 계수 즉 상관계수가 매우 작은 난점을 보완하기 위하여 신뢰도 계수를 이용한 최대 상사성계수의 값을 정의하고 그 계산 예를 제시하였다.

2. 표준화 점수들차의 확률분포인 반정규 분포와 접정규분포를 이용하여 최대상사성계수를 측정하는데 요하는 중위수 및 평가치 등급의 산정을 위한 확률하한치와 확률상한치를 구하기 위한 수표제작법 및 구체적인 예를 기술하였다.

3. 1과 2에서 구한 연구결과들을 평가자의 평가목적과 기준에 따라 효과적으로 이용하는 법과 그 평가결과의 해석에 관한 사례들을 구체적으로 기술하였다.

결론적으로 이 연구 결과를 활용하기 위한 제언으로써 우리 평가자들은 양호한 평가도구를 개발하기 위한 문제은행을 설치하고 평가자의 평가기준과 목적에 부합되는 동질, 동형의 고사를 출제하여 수험자들이 시간과 장소에 구애됨이 없이 독립적으로 그들의 진능력을 발휘할 수 있도록하고 또 여기서 논한 평가기법을 효과적으로 운용함으로써 그들의 평가 결과에 대하여 보다 더 바람직한 평가 판정을 내리도록 해야 할 것이다.

참 고 문 헌

홍석강 (1988). 신구두고사 평가치 변환에 의한 진 분포와 모수추정에 관한 연구, 한국수학교육학회지시리즈 A <수학교육> **29(2)** pp.79-93.

홍석강 (1997). 이원화평가자료의 최적평가치 산정과 평가원의 신뢰도에 관한 연구, 한국수학교육학회지시리즈 E <수학교육 프로시딩> **6** pp.159-171.

홍석강 (1999). 논문형고사 평가에서 평가치 조정과 평가원의 신뢰도 향상에 유효한 CDM 모형의 응용, 한국수학교육학회지시리즈A <수학교육> **38(2)**, pp.165-172.

Daniel, C. (1959). Use of half-normal plots in interpreting two level experiments, *Technometrics* **1**, pp.311-341.

Elandt, R. C. (1961). The Folded normal Distribution: Two methods of estimating parameters from moments. *Technometrics* **3(4)**, pp.551-562.

Jensen, A. R. (1970). IQs of identical twins reared apart. *Behavior Genetics* **1**, pp.133-148.

Leone, F. C.; Nelson, L. S. & Nottingham, R. B. (1961). The folded normal distribution, *Technometrics* **3**, pp.543-550.

McCall, R. B. (1979). *The development of intellectual functioning in infancy and the prediction of later IQ.*, In J. D. Osofsky(Ed.). Handbook of infant development, pp.707-741, N.Y.Wiley.

McGraw, K. O. & Wong S. P. (1994). The descriptive use of absolute differences between pairs of scores with common mean and variance, *Journ. of Educational Statistics* **19(2)**, pp.103-110.

Plomin, R & DeFries, J. C. (1980). *Genetics and intelligence: Recent data, Intelligence* **4**, pp.15-24.

Study on Enumerating the Degree of Similarity in Pairs of the Standardized Scores and Lower and Upper tail Probabilities using the Folded Normal Distribution.

Hong Suk-Kang
Dept. of mathematics Education, DongGuk University, 26, Pil-Dong 3- Ka, Choong-ku, Seoul, Korea, 100-715, Korea. E-mail : skhong@kra.dongguk.ac.kr

In this thesis we concerned with the degree of similarity in pairs of scores having a common mean and variance could express similarity in terms of the absolute difference between the standardized scores. We particulary discussed the distribution of absolute differences between pairs of T scores among many standardized scores and demonstrated the procedures for calculating the lower limit of Med ($|d|$) values i.e. the maximum possible similarity with medians and correlation coefficients of the equivalent form tests by using the folded normal distribution, although other researchers expressed the degree of similarity using only the standard normal distribution. We also described many cases how to use those techniques and to apply effectively them in real evaluation fields.