

Test of randomness for answers arrangement in 2017 College Scholastic Ability Test

Sojin Ahn^a · Jae Eun Lee^a · Dae-Heung Jang^{a,1}

^aDepartment of Statistics, Pukyong National University

(Received February 02, 2017; Revised April 06, 2017; Accepted April 21, 2017)

Abstract

In test with multiple choices, it is necessary to have the position of correct answers of each question spreaded evenly over all the questions in order to minimize the influence of answering tendency of test takers with preference to specific position of multiple choices. The scores of tests with correct answers in specific positions would not reflect exactly the academic aptitude of examinees who do not know correct answers but have the biased answering tendency. In this paper, we have randomness test for the positioning of correct answers at the 2017 College Scholastic Ability Test (CSAT) using Bartels rank test, the Wald-Wolfowitz runs test, the turning point test, the Cox Stuart trend test, the difference sign test and the Mann-Kendall tank test, etc. We also do independence test between the location of correct answer and the allocation of score in each question, for it may result in overestimating the test-taker with specific position preference in marking correct answers.

Keywords: College Scholastic Ability Test, test of randomness, test of independence

1. 서론

다음 글은 2016년 11월 17일 시행된 대학수학능력시험이 종료된 후 당일 인터넷신문(중앙일보)에 올라온 기사 (Kim, 2016)이다.

“17일 실시된 2017학년도 대학수학능력시험 1교시 국어영역 짝수형 객관식 문항의 정답이 유독 4번에 몰린 것으로 나타났다. 전체 45문제 중 14개(31%) 문제의 정답이 4번이라는 점에서 논란 여지가 있어 보인다. 이날 한국교육과정평가원이 1번 문항부터 4번 문항까지의 정답이 모두 4번, 5번의 답은 5번이지만 6 7번의 정답은 다시 4번이 반복된다. 정답만 나열할 경우 ‘4-4-4-4-5-4-4’가 된다. 7번까지 하나만 빼고 모두 4번이 정답이다. 4번 릴레이는 이후 15번 문제에서부터 다시 시작된다. 15 17번 문제에서 다시 4번이 반복되는 등 전체적으로 4번이 정답으로 많이 등장한다.”

Figure 1.1은 Kim (2016) 기사에 실린 2017학년도 대학수학능력시험 1교시 국어영역 짝수형 정답표를 나타낸다. 빨간 선으로 표시한 두 부분이 4번이 정답인 연속적인 문항들이다. 정답 배열이 특정 한 번호로 몰리면서 시험을 치르는 일부 수험생들이 문제의 답을 맞게 찾았어도 불안한 마음에 답을 고쳐 틀리기도 하였다. 답을 고치지 않은 학생도 시험을 치르는 동안 심적이 부담이 있었다고 한다.

¹Corresponding author: Department of Statistics, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Korea. E-mail: dhjang@pknu.ac.kr

2017학년도 대학수학능력시험

국어 영역 정답표
(짝수)형

문항 번호	정답	배점	문항 번호	정답	배점	문항 번호	정답	배점	문항 번호	정답	배점
1	㉠	2	13	㉢	2	25	㉢	3	37	㉢	2
2	㉠	2	14	㉢	2	26	㉠	2	38	㉡	2
3	㉠	2	15	㉠	2	27	㉡	2	39	㉤	3
4	㉠	2	16	㉠	2	28	㉡	2	40	㉠	2
5	㉤	3	17	㉠	2	29	㉠	2	41	㉠	2
6	㉠	2	18	㉤	3	30	㉡	2	42	㉠	2
7	㉠	3	19	㉠	2	31	㉠	3	43	㉢	2
8	㉠	2	20	㉡	2	32	㉢	2	44	㉤	2
9	㉠	2	21	㉠	2	33	㉠	2	45	㉠	3
10	㉢	2	22	㉠	3	34	㉠	2			
11	㉠	2	23	㉠	2	35	㉠	3			
12	㉤	3	24	㉢	2	36	㉢	2			

Figure 1.1. The correct answer array of 'Korean language' area in the 2017 College Scholastic Ability Test.

선다형 문항의 시험의 경우, 정답 배열과 관련해서 정답이 특정 위치에 편중되지 않고 고르게 분포되어야 한다. Sang과 Yang (2007), Han과 Kim (2014)는 오지 선다형 적성검사에서 정답이 외곽치(1, 5번 선택지)에 위치할 때보다 중앙치(2, 3, 4번 선택지)에 위치할 때 정답률이 높아진다는 실제 사례들을 제시하였다. 이러한 현상을 리커드 척도에서의 중앙집중오류(central-tendency error)라 부른다. 이러한 오류에 대한 지적은 해외 논문에서도 나타난다 (Bar-Hillel과 Attali, 2002; Attali와 Bar-Hillel, 2003, 등). 이러한 오류를 방지하기 위하여 multiple-choice test 에 대한 가이드라인이 제공되기도 한다 (Burton, 1991). 그러므로 어떤 시험의 문항의 답이 외곽치(1, 5번 선택지)보다 중앙치(2, 3, 4번 선택지)에 답이 있고 정답을 모르는 수험생이 추측하여 답을 맞힐 때 중앙치 응답경향이 있는 수험생이 외곽 문항 응답경향이 있는 수험생보다 높다고 할 수 있다. 정답이 특정 위치에 편중되지 않고 고르게 분포되어야 하는 이유는 수험생들의 응답경향이 시험에 미치는 영향을 최소화하기 위한 것으로, 특정 문항의 답지를 선호하는 응답경향이 있는 수험생에게 이익이나 불이익을 받지 않도록 하는 것이다. 즉, 이처럼 2017학년도 대학수학능력시험의 국어 짝수형 영역에 대한 논란이 있었기에, 국어 짝수형 영역을 포함한 2017학년도 대학수학능력시험의 영역별 정답 배열의 임의성을 검정하고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 임의성 검정 방법에 대하여 간략하게 소개하고 3절에서 각 영역별 정답 배열의 임의성 검정 결과를 보여준다. 영역별 정답 배열에 대하여 Bartels rank test, Wald-Wolfowitz runs test, turning point test로 음 또는 양의 상관관계가 있는지 검정하고, Cox Stuart trend test, difference sign test, Mann-Kendall rank test로 상향 또는 하향 추세가 있는지 검정한다. 음 또는 양의 상관관계 및 상향 또는 하향 추세가 있으면 정답 배열이 고르게 분포되었다고 할 수 없다. 4절에서는 각 영역별 정답과 배점에 대한 독립성 검정을 하였으며, 5절에서는 결론을 제시하였다.

2. 임의성 검정 방법

자료의 임의성을 판단하기 위한 여러 가지 다양한 통계 기법들이 있으나 본 논문에서는 다음의 통계 기법들을 선택하였다.

2.1. Bartels rank test (Bartels, 1982)

이 임의성 검정은 Bartels가 제안한 것으로 von Neumann 비검정의 비모수적 방법이다. 표본에서 i 번째 관측값의 순위를 $R_i = \text{rank}(X_i)$, $i = 1, \dots, n$ 이라 하자. 이 때의 검정통계량은 아래와 같다.

$$RVN = \frac{\sum_{i=1}^{n-1} (R_i - R_{i+1})^2}{\sum_{i=1}^{n-1} \left(R_i - \frac{n+1}{2}\right)^2}.$$

자료에 같은 값이 없으면 분모의 값은 $n(n^2 - 1)/12$ 이어서 상수가 되고, 검정통계량은 다음과 같다.

$$NM = \sum_{i=1}^{n-1} (R_i - R_{i+1})^2.$$

자료에 같은 값이 있으면 RVN의 평균과 분산은 각각 아래와 같다.

$$\mu_{RVN} = 2, \quad \sigma_{RVN}^2 = \frac{4(n-2)(5n^2 - 2n - 9)}{5n(n+1)(n-1)^2}.$$

적당한 표본 크기($10 < n < 100$)에서 $RVN/4$ 의 분포는 베타분포에 근사한다. 표본의 크기가 큰 경우에 $(RVN - \mu_{RVN})/\sigma_{RVN}$ 은 표준정규분포에 근사한다. 좌측 단측 검정을 하여 음의 상관성이 있는지 확인할 수 있다.

2.2. Cox-Stuart trend test (Cox와 Stuart, 1955)

1995년 Cox와 Stuart에 의해 소개된 이 검정은 sign test에 근거한 추세 검정이다. 뒤에 분포한 관측값들의 경향이 앞에 분포한 관측값들보다 상향 또는 하향 추세를 보여주는 지 검정한다.

$(X_1, X_{1+c}), (X_2, X_{2+c}), \dots, (X_{n-c}, X_n)$ 으로 자료를 둘 짝 짝지어 배열한다. 이 때 n 이 짝수이면 $c = n/2$ 이고, n 이 홀수이면 $c = (n+1)/2$ 이다. n 이 홀수일 경우, 중앙값은 제거한다. 그 다음은 각 짝 (X_i, X_{i+c}) 에 대해 $X_i > X_{i+c}$ 이면 “-” 부호를, $X_i < X_{i+c}$ 이면 “+” 부호를 부여한다. “+”의 갯수를 T 라고 하며, 검정통계량 T 는 시행횟수가 m 이고 확률이 $1/2$ 인 이항분포를 따른다. 여기서 m 은 둘 짝 짝지어진 배열의 개수이다. 양측검정 시행으로 상향 또는 하향 추세를 알 수 있으며, T 값이 큰 경우 상향 추세가 있다고 추정한다.

2.3. Difference sign test (Moore와 Wallis, 1943)

연속적인 사건들 X_i , $i = 1, \dots, n$ 에서 $X_i > X_{i-1}$ 인 점의 개수를 센다. 이 때 통계량 S 를 다음과 같이 정의한다.

$$S = \sum_{i=2}^n Y_i.$$

이 때 $X_i > X_{i-1}$ 이면 $Y_i = 1$ 이고 나머지의 경우 $Y_i = 0$ 이다. 평균은 $\mu_s = (n-1)/2$ 이고, 분산은 $\sigma_s^2 = (n+1)/12$ 이다. 상향 또는 하향 추세가 있는지 양측 검정을 통해 확인하며, 단측 검정 시행시 S 가 클 경우 상향 추세로 추정한다.

2.4. Mann-Kendall rank test (Mann, 1945; Kendall, 1990)

Mann이 제안한 이 검정은 $i < j$ 인 모든 (X_i, X_j) 에 대하여 추세를 비교한다. 통계량 P 는 $X_i < X_j$ 인 짝의 개수이다.

$$P = \sum_{1 \leq i < j \leq n} Y_{ij}, \quad Y_{ij} = \begin{cases} 1, & \text{if } X_i < X_j, \\ 0, & \text{otherwise.} \end{cases}$$

평균과 분산은 각각 $\mu_P = n(n-1)/4$, $\sigma_P^2 = n(n-1)(2n+5)/72$ 이며, P 값이 크면 상향 추세를 보인다.

2.5. Wald-Wolfowitz runs test (Wald와 Wolfowitz, 1940)

Wald와 Wolfowitz가 처음 제안한 것으로 같은 모집단에서 추출된 두 독립 표본을 검정한다. 임의의 배열에서 R 갯수를 센다. 여기서 R 은 연(run)의 수를 말한다. 예를 들어, 표본 (A, A, B, A, B, B, A) 에서 $R = 5$ 이다. 귀무가설 하의 분포는 다음과 같다.

$$P(R=r) = \begin{cases} \frac{2 \binom{n_1-1}{r/2-1} \binom{n_2-1}{r/2-1}}{\binom{n_1+n_2}{n_1}}, & \text{if } r = \text{even}, \\ \frac{\binom{n_1-1}{(r-1)/2} \binom{n_2-1}{(r-3)/2} + \binom{n_1-1}{(r-3)/2} \binom{n_2-1}{(r-1)/2}}{\binom{n_1+n_2}{n_1}}, & \text{if } r = \text{odd}, \end{cases}$$

여기서 $r = 2, 3, \dots, 2 \times \min(n_1, n_2) + c$ 이다. 단 n_1 과 n_2 가 같으면 $c = 0$ 이고, 같지 않으면 $c = 1$ 이다.

$$\mu_R = 1 + \frac{2n_1n_2}{n_1+n_2}, \quad \sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

n_1 과 n_2 의 표본의 개수가 각각 20보다 클 때는 $(R - \mu_R)/\sigma_R$ 은 표준정규분포에 근사한다. 양측검정에서 R 이 매우 작거나 클 때는 귀무가설을 기각하며, 우측 단측검정에서 R 이 크면 음의 상관성이 있다고 할 수 있다.

2.6. Turning point test (Moore와 Wallis, 1943; Brockwell과 Davis, 2002)

$i = 1, \dots, n$ 인 연속적인 임의의 사건들 X_i 가 있을 때, (X_{i-1}, X_i, X_{i+1}) 은 6개 가능한 순서로 동등하게 발생할 가능성이 있다. 그 중 4개는 전환점(turning point)이 된다. 중앙에 세 점 중의 가장 큰 값 또는 작은 값 X_i ($2 < i < n-1$)가 있을 때, $X_{i-1} < X_i$ 이고 $X_i > X_{i+1}$ 이거나 $X_{i-1} > X_i$ 이고 $X_i < X_{i+1}$ 이라 하면 X_i 는 전환점이 된다. 전환점의 확률은 $2/3$ 이다. 전환점의 갯수를 검정통계량 T 로 나타낸다. 이 때의 평균과 분산은 각각 $\mu_T = 2(n-2)/3$, $\sigma_T^2 = (16n-29)/90$ 이다. 표본의 크기가 클 때, $(T - \mu_T)/\sigma_T$ 는 표준정규분포에 근사한다. 좌측 단측검정으로 인접한 관측값 사이의 양의 상관관계가 있는지 확인할 수 있다.

3. 영역별 정답 배열의 임의성 검정

크게 3 영역(1. 국어, 수학, 영어, 한국사 영역의 짝수형, 홀수형; 2. 사회탐구 영역; 3. 과학탐구 영역)으로 나누어 임의성 검정을 하였다.

Table 3.1. *P*-values of tests of randomness for 'Korean', 'Math', 'English', and 'Korean History' area

과목	Bartels rank test		Cox-Stuart trend test		Difference sign test		Mann-Kendall rank test		Wald-Wolfowitz runs test		Turning point test	
	Two-sided		Two-sided		Two-sided		Two-sided		Two-sided		Two-sided	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
국어	0.875		0.167		0.553		0.000		0.318		0.066	
짝수형	-	-	0.084	0.968	-	-	0.000	1.000	-	-	0.967	0.033
국어	0.803		0.332		0.574		0.009		0.900		0.505	
홀수형	-	-	-	-	-	-	0.005	0.995	-	-	-	-
수학 가형	0.610		0.754		0.233		0.165		0.813		0.049	
짝수형	-	-	-	-	-	-	0.082	0.918	-	-	0.975	0.025
수학 가형	0.724		0.688		0.691		0.673		0.645		0.169	
홀수형	-	-	-	-	-	-	-	-	-	-	0.916	0.084
수학 나형	0.667		1.000		0.257		0.904		0.790		0.578	
짝수형	-	-	-	-	-	-	-	-	-	-	-	-
수학 나형	0.256		0.727		0.706		0.856		0.203		0.095	
홀수형	-	-	-	-	-	-	-	-	-	-	0.952	0.048
영어	0.568		1.000		0.417		0.257		0.620		0.609	
짝수형	-	-	-	-	-	-	-	-	-	-	-	-
영어	0.375		0.815		1.000		0.109		0.603		0.130	
홀수형	-	-	-	-	-	-	0.054	0.946	-	-	0.935	0.065
한국사	0.904		0.688		0.233		0.382		0.173		0.326	
짝수형	-	-	-	-	-	-	-	-	-	-	-	-
한국사	0.302		0.039		0.414		0.807		0.409		0.543	
홀수형	-	-	0.998	0.020	-	-	-	-	-	-	-	-

3.1. 국어, 수학, 영어, 한국사 영역

Table 3.1은 ‘국어’, ‘수학’, ‘영어’, ‘한국사’ 영역의 정답 배열에 임의성이 있는지 여러 가지 통계 기법으로 검정한 결과이다. 양측검정 시행 후 임의성이 없을 경우 단측검정을 시행하여 추세를 알아보았다. 어떤 추세가 있는지 알기위하여 시행한 단측검정에서 유의수준 0.1로 검정을 하였기에 양측검정은 유의수준 0.2를 기준으로 검정하였다. ‘국어 짝수형’에서 임의성이 없다는 결과가 나오는 통계기법은 Cox Stuart trend test, Mann-Kendall rank test, turning point test이다. 특히 Mann-Kendall rank test 결과는 *p*값이 0으로 임의성이 지켜지지 않은 것으로 보이며, 단측검정에서 강한 하향 추세를 보인다. Cox Stuart trend 좌측 단측검정에서 약한 감소추세를 확인했으며, turning point 우측 단측검정에서 음의 상관관계를 확인할 수 있다. 그 밖에 ‘국어 홀수형’, ‘수학 가형 짝수형’과 ‘수학 가형 홀수형’, ‘영어 홀수형’의 경우에는 Mann-Kendall rank test에서 임의성이 없다는 결과를 보이며, 좌측 단측검정 결과에서 강한 하향추세를 보인다. 또한 turning point test에서 ‘수학 가형 짝수형’과 ‘수학 가형 홀수형’, ‘수학 나형 홀수형’, ‘영어 홀수형’의 임의성이 지켜지지 않으며, 우측 단측검정 결과에서 음의 상관관계를 알 수 있다.

Figure 3.1은 ‘국어 짝수형’, ‘사탐 세계사’, 그리고 ‘과탐 생명과학1’의 정답 배열의 산점도이다. Figure 3.1의 각 산점도에서 점을 선으로 이어 연속된 변동을 알 수 있도록 하였다. 또한 산점도를 4개의 영역으로 나누어 정답이 한 부분에 집중되었는지 확인한다. Figure 3.1 ‘국어 짝수형’ 산점도에서 제2사분면에 큰 변동이 없으며 4번 정답이 시험 앞 쪽에 집중적으로 나타난 것을 알 수 있다. 결과를 종합하였을 때 다른 영역에 비하여 ‘국어 짝수형’에서 정답 배열의 임의성이 지켜지지 않았다고 할 수 있다.

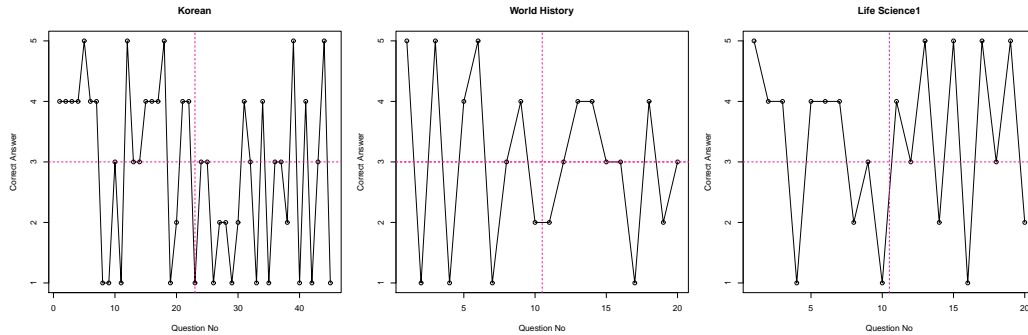


Figure 3.1. Scatter plot of answer arrangement for 'Korean', 'World History', and 'Life Science1'.

Table 3.2. P-values of tests of randomness for 'Social' area

과목	Bartels		Cox-Stuart		Difference		Mann-Kendall		Wald-Wolfowitz		Turning	
	rank test		trend test		sign test		rank test		runs test		point test	
	Two-sided		Two-sided		Two-sided		Two-sided		Two-sided		Two-sided	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
생활과	0.148		0.344		0.414		0.948		0.184		0.224	
윤리	0.074	0.926	-	-	-	-	-	-	0.092	0.926	-	-
윤리와	0.559		0.070		0.655		0.270		0.291		0.042	
사상	-	-	-	-	-	-	-	-	-	-	0.979	0.021
한국지리	0.296		1.000		0.414		0.650		0.121		0.543	
	-	-	-	-	-	-	-	-	0.940	0.060	-	-
세계지리	0.147		0.688		0.414		0.746		0.338		0.068	
	0.073	0.927	-	-	-	-	-	-	-	-	0.034	0.966
동아시아사	0.927		1.000		0.691		0.218		0.263		0.326	
	-	-	-	-	-	-	-	-	-	-	-	-
세계사	0.054		1.000		0.414		0.092		0.057		0.543	
	0.973	0.027	-	-	-	-	0.046	0.954	0.972	0.029	-	-
법과정치	0.224		0.688		0.121		0.399		0.301		0.703	
	-	-	-	-	-	-	-	-	-	-	-	-
경제	0.911		1.000		0.674		0.069		0.442		0.294	
	-	-	-	-	-	-	0.035	0.965	-	-	-	-
사회·문화	0.769		0.508		1.000		0.038		0.646		0.029	
	-	-	-	-	-	-	0.019	0.981	-	-	0.985	0.015

3.2. 사회탐구 영역

Table 3.2는 '사회탐구' 영역 정답 배열에 대한 임의성의 양측 및 단측 검정 결과이다. 다른 검정에 비하여 Bartels rank test, Mann-Kendall rank test, Wald-Wolfowitz runs test, turning point test 에서 임의성이 없다는 결과가 자주 보인다. 각 기법에서 임의성이 없다고 판단되는 영역에 대하여 추가로 단측검정을 실시하였다. Mann-Kendall rank 단측검정에서 '세계사', '경제', '사회·문화' 영역에서 강한 하강추세를 알 수 있다. turning point 단측검정에서 '윤리와 사상'과 '사회·문화'에서는 음의 상관관계를, '세계지리' 영역에서는 양의 상관관계를 확인할 수 있다. 다른 영역에 비하여 '세계사' 영역에서 3개 통계기법에서 임의성이 지켜지지 않은 결과를 보여준다.

Table 3.3. P-values of tests of randomness for 'Science' area

과목	Bartels		Cox-Stuart		Difference		Mann-Kendall		Wald-Wolfowitz		Turning	
	rank test		trend test		sign test		rank test		runs test		point test	
	Two-sided		Two-sided		Two-sided		Two-sided		Two-sided		Two-sided	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
물리1	0.292		1.000		1.000		0.846		1.000		0.224	
	-	-	-	-	-	-	-	-	-	-	-	-
물리2	0.955		1.000		0.674		0.196		0.774		0.834	
	-	-	-	-	-	-	-	-	-	-	-	-
화학1	0.508		0.180		1.000		0.006		0.184		0.068	
	-	-	-	-	-	-	0.003	0.997	0.092	0.908	0.966	0.034
화학2	0.499		0.508		0.691		0.897		0.168		0.011	
	-	-	-	-	-	-	-	-	0.916	0.084	0.995	0.005
생명과학1	0.048		0.344		0.414		0.153		0.117		0.015	
	-	-	-	-	-	-	0.077	0.923	0.942	0.058	0.993	0.007
생명과학2	0.332		1.000		0.233		0.330		0.605		0.432	
	-	-	-	-	-	-	-	-	-	-	-	-
지구과학1	0.497		0.508		0.674		0.092		0.324		0.675	
	-	-	-	-	-	-	0.046	0.954	-	-	-	-
지구과학2	0.518		1.000		0.103		0.270		0.605		0.224	
	-	-	-	-	-	-	-	-	-	-	-	-

Figure 3.1의 '사탐 세계사' 산점도에서 1사분면과 2사분면의 1번과 5번의 반복으로 변동이 큰 반면 3사분면과 4사분면은 변동이 비교적 작으며 4번이나 5번 정답 후 1번이나 2번 정답이 나오는 경향이 반복되어 하강추세를 확인할 수 있다.

3.3. 과학탐구 영역

Table 3.3은 과학탐구 영역 정답 배열에 대한 임의성의 양측 및 단측 검정 결과이다. Mann-Kendall rank test 결과에서 '화학1', '생명과학1', '지구과학1' 영역은 하강추세가 있으며, turning point test 결과에서 '화학1', '화학2', '생명과학1' 영역은 음의 상관관계가 보인다. Wald-Wolfowitz runs test 결과에서 '화학2'와 '생명과학1' 영역은 음의 상관관계를, '화학1' 영역은 양의 상관관계를 보인다. Figure 3.1의 '과탐 생명과학1' 산점도에서 4번이나 5번 정답 후 1번이나 2번 정답이 나오는 경향이 반복되어 음의 상관관계 및 하강추세를 확인할 수 있다.

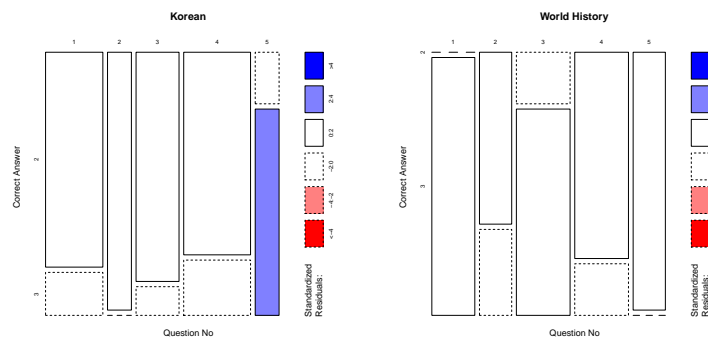
4. 정답 문항과 배점에 대한 독립성 검정

Table 4.1은 각 영역의 정답과 배점 관계 독립성 여부를 검정한 결과이다. 검정 결과 '국어영역 짝수형'의 정답 문항과 배점이 종속 관계임을 알 수 있다. 정답 문항과 배점 사이의 관계에서 독립성이 지켜지지 않으면 앞 절에서 언급하였듯이 정답을 모르는 수험생이 추측하여 문항의 답을 맞혔을 때 높은 점수 또는 낮은 점수를 받을 확률이 높아 응답경향이 시험 결과에 영향을 미친다.

특정 각 정답별 배점 비율을 보여주기 위하여 모자이크그림을 그렸다. Figure 4.1에서 '국어영역 짝수형'은 2번이 정답인 문항은 3점 배점이 하나도 없었고, 5번이 정답인 문항은 3점 배점이 상대적으로 많다. 다른 문항보다 3점 배점이 많은 5번 문항의 표준화 잔차값이 커서 파란 색깔로 표시되었다. 모자

Table 4.1. *P*-values of tests of independence

과목	카이제곱 검정	Fisher의 정확검정
국어 짝수형	0.018	0.037
국어 홀수형	0.732	0.806
수학 가형 짝수형	0.427	0.690
수학 가형 홀수형	0.933	0.971
수학 나형 짝수형	0.789	0.886
수학 나형 홀수형	0.800	0.883
영어 짝수형	0.530	0.605
영어 홀수형	0.440	0.455
한국사 짝수형	0.121	0.146
한국사 홀수형	0.178	0.241
생활과 윤리	0.900	1.000
윤리와 사상	0.363	0.476
한국 지리	0.087	0.107
세계 지리	0.099	0.120
동아시아사	0.639	0.790
세계사	0.027	0.021
법과 정치	0.323	0.497
경제	0.061	0.082
사회·문화	0.736	0.813
물리1	0.316	0.425
물리2	0.864	0.958
화학1	0.639	0.790
화학2	0.443	0.610
생명과학1	0.424	0.496
생명과학2	0.821	0.930
지구과학1	0.147	0.217
지구과학2	0.639	0.790

**Figure 4.1.** Mosaic plot between the location of correct answer and the allocation of score in each question for 'Korean' and 'World History'.

이크그림을 통해서도 '국어영역 짝수형'의 정답 문항과 배점이 종속 관계임을 알 수 있다. 그 외 영역 중에서 '사회탐구 세계사' 영역의 정답과 배점 관계 독립성이 지켜지지 않은 결과가 나왔지만 Figure

4.1에서 볼 수 있듯이 그 정도가 국어영역 짝수형에 비해 심각하지 않은 것을 알 수 있다. 다른 영역의 독립성에는 문제가 없으므로 모자이크그램을 제시하지 않았다.

5. 결론

본 연구에서는 2017학년도 대학수학능력시험의 각 영역별 정답 배열에 대한 임의성 검정 및 정답과 배점에 대한 독립성 검정을 실시하였다. 임의성 검정 결과 다른 영역에 비해 ‘국어영역의 짝수형’ 영역에서 임의성이 강하게 지켜지지 않은 결과를 보여주었다. 정답과 배점에 대한 독립성 검정에서는 ‘국어영역 짝수형’ 영역과 ‘사회탐구 세계사’ 영역에서 정답과 배점이 종속관계로 나타났다. 분석 결과를 종합하여 볼 때 다른 영역에 비하여 ‘국어영역 짝수형’이 정답 배열의 임의성 및 정답과 배점의 독립성이 지켜지지 않아 수험생의 응답경향이 시험 결과에 영향을 미치므로 문제가 된다. 이 결과를 교훈삼아 추후 시행되는 대학수학능력시험에서는 정답 배열의 임의성 및 정답과 배점의 독립성을 지키는 데 만전을 기하여야 할 것이다.

References

- Attali, Y. and Bar-Hillel M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable, *Journal of Educational Measurement*, **40**, 109–128.
- Bar-Hillel, M. and Attali, Y. (2002). Seek whence: answer sequences and their consequences in key-balanced multiple-choice tests, *The American Statistician*, **56**, 299–303.
- Bartels, R. (1982). The rank version of von Neumann’s ratio test for randomness, *Journal of the American Statistical Association*, **77**, 40–46.
- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting* (2nd ed.), Springer, New York.
- Burton, S. J. (1991). How to prepare better multiple-choice test items: Guidelines for university faculty, *Brigham Young University Testing Services and The Department of Instructional Science*.
- Cox, D. R. and Stuart, A. (1955). Some quick sign test for trend in location and dispersion, *Biometrika*, **42**, 80–95.
- Han, Y. S. and Kim, M. S. (2014). A study on the item characteristics differences of response position, response length, and question types of multiple-choice aptitude tests, *Journal of the Korea Academia-Cooperation Society*, **15**, 3609–3615.
- Kendall, M. (1990). *Rank Correlation Methods* (5th ed.), Oxford University Press, Oxford.
- Kim, Y. M. (2016). Examinees are just completely blown because of ‘4444544’... correct answer relay during Korean language area test in College Scholastic Ability Test, JoongAng Daily, from: <http://news.joins.com/article/20889947>.
- Mann, H. B. (1945). Nonparametric test against trend, *Econometrica*, **13**, 245–259.
- Moore, G. H. and Wallis, W. A. (1943). Time series tests based on sign of differences, *Journal of the American Statistical Association*, **38**, 153–154.
- Sang, K. A. and Yang, K. S. (2007). The effect of response position in a multiple-choice examination, *Asian Journal of Education*, **8**, 25–46.
- Wald, A. and Wolfowitz, J. (1940). on a test whether two samples are from the same population, *The Annals of Mathematical Statistics*, **11**, 147–162

2017학년도 대학수학능력시험 영역별 정답배열 임의성 검정

안소진^a · 이재은^a · 장대흥^{a,1}

^a부경대학교 통계학과

(2017년 02월 02일 접수, 2017년 04월 06일 수정, 2017년 04월 21일 채택)

요약

선다형 문제의 경우 특정 위치의 답을 선호하는 수험생의 응답경향이 시험 결과에 미치는 영향을 최소화하기 위하여 정답배열을 특정위치에 편중되지 않고 고르게 분포할 필요가 있다. 이는 정답배열이 특정 번호에 편중되어 있으면 정답을 모를 때 중앙 또는 외곽 번호에 응답경향이 있는 수험생이 추측으로 문항의 답을 맞힐 수 있으므로 시험 점수가 학생의 능력을 제대로 반영하지 못하는 결과를 초래할 수 있기 때문이다. 본 논문에서는 2017학년도 대학수학능력시험의 영역별 정답배열에 대하여 Bartels rank test, Wald-Wolfowitz runs test, turning point test, Cox Stuart trend test, difference sign test, Mann-Kendall rank test 등을 실시하여 정답배열이 고르게 분포되었는지 확인하였다. 또한 정답 번호와 배점 사이에 독립성이 지켜지지 않을 경우에도 특정 번호를 정답으로 선호하는 경향을 가진 수험생의 점수는 학생의 능력을 제대로 반영하지 못할 수 있기 때문에 정답의 위치와 배점 사이의 독립성 검정을 추가로 실시하였다.

주요용어: 2017학년도 대학수학능력시험, 임의성 검정, 독립성 검정

¹교신저자: (48513) 부산광역시 남구 용소로 45, 부경대학교 통계학과. E-mail: dhjang@pknu.ac.kr