

사영에 근거한 면접 점수의 통계적 모형

박철용¹ · 김현욱²

¹²계명대학교 통계학과

접수 2012년 4월 22일, 수정 2012년 5월 5일, 게재확정 2012년 5월 15일

요약

이 연구에서는 면접 점수가 발생하는 통계적 모형을 사영에 근거하여 제시하였다. 이 모형은 피면접자의 개인별 참값과 이와 관련된 변수값이 2차원 평면의 X와 Y축의 값으로 주어졌을 때, 심사위원의 시각을 X축과의 각도로 생각하여 이 축에 사영된 값으로서 심사위원의 면접 점수의 평균으로 잡는 방법이다. 이 값에 개인적 편향과 관측 오차를 더해져 심사위원의 관측 면접 점수가 얻어지게 된다. 이 통계적 모형을 사용하여 흔히 사용되고 있는 면접 점수 표준화 방법인 절사평균법, 순위평균법, z-점수평균법을 비교하였다. 이 모의실험에서는 두 가지 면접 형태, 두 가지 면접자 수, 두 가지 면접자의 전문성 정도, 실제 점수와 관련된 변수 간의 분포 두 가지와 세 가지 상관계수가 고려되었다.

주요용어: 사영, 순위상관, 순위평균, 절사평균, 표준점수평균.

1. 서론

대학입시와 입사시험에서 지필고사보다 면접이 차지하는 비중이 확대되는 추세이다. 이는 지필고사를 통해 평가하기 어려운 인성, 적성 및 도전정신 등을 평가하기에 면접이 보다 적합하기 때문이다. 이러한 면접 강화 추세는 수험생과 입사준비생에게 상당한 부담을 안겨주고 있지만 대학과 회사 입장에서도 유능한 수험생과 입사준비생들을 유치하기 위해서 객관적인 면접점수 표준화 방법을 도입하여야 하는 부담이 생겼다.

수능시험에서 선택과목 간의 난이도 차이를 극복하기 위해서 표준점수 제도가 도입되면서, 대학입시에서 객관성을 보장하기 위한 방법에 대한 연구가 본격적으로 진행되었다. 허명희 (1994)에서는 표준점수제에서 유용한 교육측정의 기본 이론이 소개되고 있으며, 박성현 등 (2000)에서는 대학입시에서 선택과목점수 표준화에 있어 등분위수화가 선택과목 점수의 변별력을 증가시키는 효과가 있다고 하였다. 황형태 (2005)에서는 선택과목별 가산점수제를 통해 표준점수제의 보완이 가능하다고 제안하였다. 기타 표준화 방법에 대한 연구로는 민대기와 정지현 (2012)과 이재만과 차영준 (2012) 등이 있다.

수능시험에서도 사용되고 있듯이 표준화 방법으로 가장 많이 사용되는 것이 z-점수 (z score)에 근거한 방법이다. 황형태 등 (2004)에서는 심사자 별 순위 (rank)에 근거한 면접점수 표준화 방법이 제안되었다. 이 방법은 심사자별 점수에 순위를 매기고 그 순위에 해당하는 표준정규분포에 근거한

¹ 교신저자: (704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교 통계학과. 교수.
E-mail: cypark1@kmu.ac.kr

² (704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교 통계학과, 석사.

점수를 부여하는 방식으로, 상위 5%와 하위 5%에는 동일한 점수를 부여하여 강건성 (robustness)을 증대시켰다.

순위에 근거한 이러한 방법과 더불어, 여러 명의 심판이 심사하는 다이빙과 체조 같은 스포츠 종목에서는 절사평균 (trimmed mean)에 근거한 표준화 방법이 많이 사용되고 있다. 예를 들어 5명의 심판이 심사하는 경우 최하와 최고의 점수를 버리고 중간 3명의 점수를 평균하여 사용하는 것이다.

이 연구에서는 면접 점수가 발생하는 통계적 모형을 사영 (projection)에 근거하여 제시하고자 한다. 구체적으로 피면접자의 개인별 참값과 이와 관련된 변수값이 2차원 평면의 X와 Y축의 값으로 주어졌을 때, 심사위원의 시각을 X축과의 각도로 생각하여 이 축에 사영된 값으로서 심사위원의 면접 점수 평균을 잡는 방법이다. 이 값에 개인적 편향과 관측 오차를 더해져 심사위원의 관측 면접 점수가 얻어지게 된다.

이 통계적 모형은 박철용 (2011)에 제시된 가중평균에 의한 통계적 모형을 발전시킨 것이다. 우선 면접 점수가 얻어지는 과정에 기하학적 의미를 가미하였다는 점이 달라졌다. 다시 말해 심사위원의 시각을 참값과의 각도로 생각하였고, 면접점수는 이 각도를 가지는 축에 사영된 참값으로 잡은 것이다. 다음으로 앞의 연구에서는 X와 Y축의 값을 참값과 잡음변수이라 명명하고 서로 독립인 상황을 가정한 반면, 이 연구에서는 참값과 관련 변수라 명명하고 두 변수 사이의 상관관계가 존재하는 경우까지 확장한 점이 달라졌다.

이 연구에서는 모의실험을 통해 세 가지 표준화 방법, 즉 절사평균, 순위평균 및 z-점수평균에 근거한 방법들을 비교하고자 한다. 구체적으로 이 모의실험에서는 한 면접자가 모든 피면접자를 심사하는 전체면접 형태와 한 면접자가 피면접자의 절반만 심사하는 반분면접을 고려한다. 또한 면접자의 수가 3명인 경우와 5명인 경우를 고려하며, 면접자가 전문가 집단인 경우와 비전문가 집단인 경우를 고려한다. 마지막으로 참값과 관련된 관련 변수의 분포로는 정규분포와 지수분포가 고려되며, 참값과 관련 변수간의 상관계수가 없을 때와 0.4, 0.8인 경우를 고려한다.

이 연구에서 사용하는 순위평균은 각 심사자의 피면접자들 점수에 순위를 매긴 후 심사자들의 순위평균을 사용하는 것으로 황형태 등 (2004)에서 사용하는 방법과 약간의 차이가 있다. z-점수평균은 각 심사자의 피면접자들 점수의 z-점수를 계산한 후 심사자들의 z-점수를 평균한 것이다. 다시 말해 세 가지 방법에 의해 관측되는 심사자의 점수는 절사평균을 제외하고는 피면접자의 참값과 다른 척도를 가지게 되는 것이다. 따라서 이 모의실험에서는 Kendall (1928)의 순위상관계수를 사용하여 세 가지 방법이 얼마나 충실히 피면접자의 참값의 순서를 유지하는지 비교하게 된다. 순위상관계수를 사용함으로써 해서 부가적으로 z-점수평균과 순위평균에 어떠한 단조증가함수 (monotone increasing function)에 의한 변환을 하여도 동일한 모의실험 결과를 얻을 수 있는 것이다.

이 논문은 다음과 같이 구성되어 있다. 2절에서는 이 논문에서 제안하고 있는 사영에 의한 면접 점수의 통계적 모형과 모의실험 결과를 제시한다. 3절에서는 이 연구의 결론과 한계점 및 추후연구과제에 대해서 논의한다.

2. 통계적 모형과 모의실험

이 연구에서는 면접 점수가 발생하는 통계적 모형을 사영 (projection)에 근거하여 제시하고자 한다. 구체적으로 피면접자의 개인별 참값과 이와 관련된 변수값이 2차원 평면의 X와 Y축의 값으로 주어졌을 때, 심사위원의 시각을 X축과의 각도로 생각하여 이 축에 사영된 값으로서 심사위원의 면접 점수의 평균으로 잡는 방법이다. 이 값에 개인적 편향과 관측 오차를 더해져 심사위원의 관측 면접 점수가 얻어지게 된다.

이 통계적 모형을 사용하여 흔히 사용되고 있는 면접 점수 표준화 방법인 절사평균법, 순위법, z-점

수법을 비교하게 된다. 이 모의실험에서는 한 면접자가 모든 피면접자를 심사하는 전체면접 형태와 한 면접자가 피면접자의 절반만 심사하는 반분면접이 고려된다. 또한 면접자의 수가 3명인 경우와 5명인 경우가 고려되며, 면접자가 전문가 집단인 경우와 비전문가 집단인 경우가 고려된다. 마지막으로 참값과 관련된 관련 변수의 분포로는 정규분포와 지수분포가 고려되며, 참값과 관련 변수 간에 상관계수가 없을 때와 0.4, 0.8인 경우가 고려된다.

2.1. 사영에 의한 면접 점수의 통계적 모형

면접자의 시각을 θ , 피면접자의 참값을 X , 관련된 변수를 W , 면접자의 성향편의를 B 그리고 피면접자 관측오차를 ϵ 이라 표기하자. 면접자의 시각 θ , 피면접자의 참값 X , 관련된 변수 W 및 성향편의 B 가 주어졌을 때 면접자의 평가점수 Y 는 다음과 같이 관측된다고 가정한다.

$$Y = \cos(\theta)X + \sin(\theta)W + B + \epsilon \quad (2.1)$$

여기서 $\cos(\theta)X + \sin(\theta)W$ 부분은 이 연구에서 제안하는 핵심적인 부분으로서 이것은 이차원의 (X, W) 값을 X 축을 θ 만큼 회전시킨 축에 사영 (projection)한 값이다 (Johnson과 Wichern, 1992, p. 43). 구체적으로 $\cos(\theta)X + \sin(\theta)W$ 부분을 그림으로 나타낸 것이 그림 2.1에 주어져 있다.

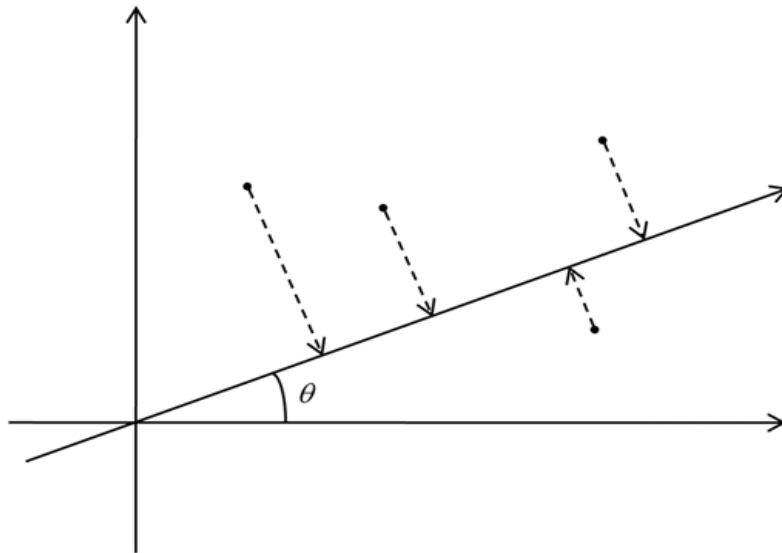


그림 2.1 면접자의 시각이 θ 인 경우 사영에 의해 면접 점수가 구해지는 과정

이 모형에서 θ 는 피면접자를 평가하는 면접자의 시각이 되는데, 0에 가까울수록 참값에 가깝게 정확하게 평가하며 0에서 벗어날수록 참값을 제대로 평가하지 못하게 된다. 면접자의 성향편의 B 는 후하게 (혹은 인색하게) 점수를 매기는 성향이 있는 사람은 모든 피면접자들에게 동일하게 후하게 (혹은 인색하게) 점수를 매기는 점을 고려하여 모형에 포함시켰다. 사영 점수에 성향편의 B 와 관측오차 ϵ 가 더해져 최종 관측 면접 점수 Y 가 얻어지는 것이다.

2.2. 모의실험

이 연구에서 고려하는 모의실험의 구조는 다음과 같다. X 와 W 의 분포로는 표준정규분포와 양의 왜도 (skewness)가 심한 지수분포를 고려한다. 일반적으로 z -점수는 정규분포에서 좋은 성능을 발휘하고 순위는 왜도가 심하거나 꼬리가 두꺼운 분포에서 좋은 성능을 보이리라 기대하기 때문에 이 두 분포를 고려한 것이다. X 와 W 는 서로 상관을 맺을 수도 있을 것이기 때문에 두 변수의 상관계수가 0, 0.4, 0.8인 경우를 고려하였다. θ 의 분포는 기본적으로 균일분포를 사용하였다. 면접자가 $\pm 25^\circ$ 사이의 균일분포를 따를 경우 전문가 집단, $\pm 50^\circ$ 사이에 속하되 $\pm 25^\circ$ 사이에 속할 가능성보다 그 밖의 범위에 속할 가능성이 2배 높은 혼합 (mixed) 균일분포를 따를 경우 비전문가 집단이라 명명하였다. 면접자가 모든 피면접자를 심사하지 않는 면접에서는 아주 중요하게 작용할 것이라 기대되는 면접자의 성향편의 B 의 분포로는 평균이 0인 정규분포를 사용하였다. 피면접자를 대하는 면접자의 관측오차 ϵ 의 분포로도 통상적으로 사용되는 평균이 0인 정규분포를 사용하였다. X 와 W 분산은 1로 고정시켰으며 B 와 ϵ 의 표준편차는 각각 1/2과 1/3로 잡아 체계적인 변동부분의 1/2, 1/3에 해당되도록 하였다.

면접자 $i = 1, 2, \dots, p$ 가 피면접자 $j = 1, 2, \dots, m$ 의 평가점수를 부과하는 과정을 요약하면 다음과 같다. 참값 X_j 를 가진 피면접자가 θ_i ($i = 1, \dots, p$)의 시각과 B_i ($i = 1, \dots, p$)의 성향편의를 가지는 면접자 그룹에 들어오면 면접자들은

$$Y_{ij} = \cos(\theta_i)X_j + \sin(\theta_i)W_j + B_i + \epsilon_{ij} \quad (2.2)$$

로 심사점수를 매기게 된다. 이 심사점수는 평균적으로 $\cos(\theta_i)X_j + \sin(\theta_i)W_j + B_i$ 이며 ϵ_{ij} 만큼의 관측오차를 동반하게 된다. 안정적인 모의실험 결과를 위해서는 m 이 크면 좋지만, 한 면접자가 응대할 수 있는 피면접자의 수는 현실적으로 한계가 있기 때문에 $m = 50$ (반분면접) 혹은 $m = 100$ (전체면접)으로 고정하였다. 따라서 전체 피면접자의 수는 $n = 100$ 으로 고정되었다.

구체적인 모의실험 설계를 변인별로 정리하면 다음과 같다. 세 가지 방법을 비교하기 위해 먼저 X 와 W 의 상관계수가 0, 0.4, 0.8인 경우로 나누어 비교한다. 다음으로 면접자가 피면접자 모두 면접하는 전체면접과 면접자가 두 그룹으로 나뉘어 피면접자의 절반만 면접하는 반분면접으로 나누어 비교한다. 구체적으로 전체면접 시 필요한 면접자는 p 명이고 이들이 100명을 모두 면접하게 되며, 반분면접 시는 $2p$ 명의 면접자가 두 그룹으로 나뉘어 각 그룹별로 50명씩 면접하게 된다. 그리고 면접자의 수 p 가 3 혹은 5가 될 때 세 가지 방법을 비교한다. 이 인원수는 절사평균이 적용될 수 있는 현실적인 면접자의 수라고 생각되어 사용되었다. 이와 더불어 θ 가 $\pm 25^\circ$ 사이의 균일분포를 따르는 전문가 집단과 $\pm 50^\circ$ 사이의 혼합 균일분포를 따르는 비전문가 집단에 따른 비교와 함께, X 와 W 의 분포가 표준정규분포와 지수분포로 변동할 때의 비교를 시도한다. 여기서 사용하는 지수분포는 정규분포와 평균과 분산이 같도록 만들기 위해서 $T - 1$ (T 는 평균이 1인 지수확률변수) 형태를 사용한다.

세 가지 방법에서 최종 면접점수를 얻는 과정은 다음과 같다. 절사평균법은 Y_{ij} ($i = 1, 2, \dots, p$) 중에서 최소값과 최대값을 제외한 표본평균으로 최종 면접점수를 정한다. 순위평균법은 Y_{ij} ($j = 1, 2, \dots, m$) 중에서의 순위 R_{ij} 를 구한 후 면접자들의 순위평균 $\sum_i R_{ij}/p$ 로서 최종 면접점수를 정한다. 또한 z -점수평균법은 Y_{ij} ($j = 1, 2, \dots, m$)의 z -점수 Z_{ij} 를 구한 후 면접자들의 z -점수 평균 $\sum_i Z_{ij}/p$ 로서 최종 면접점수를 정한다.

세 가지 방법의 성능을 비교하는 기준은 Kendall (1938)의 순위상관계수이다. 구체적으로 참값 X 와 세 가지 방법의 최종 면접점수 간의 순위상관계수를 계산하여 이 값이 가장 큰 방법을 최고의 성능을 가지는 방법으로 인정하게 된다. 1000번의 반복실험을 통해 참값과 최종 면접점수 간의 순위상관계수를 계산하여 표본평균과 표준오차를 계산하여 정리한 것이 표 2.1에서 표 2.3에 나타나 있다. 모의실험을 위해 R과 Minitab 16이 사용되었다.

표 2.1 실제 점수와 세 가지 표준화 방법들 간의 순위상관계수 (상관계수=0)

| 면접유형 | 면접자수 | 분포 | 면접자집단 | 표준화 방법 | | |
|------|------|---------------|---------------|---------------|---------------|---------------|
| | | | | 점사평균 | 순위평균 | z-점수평균 |
| 전체면접 | 3 | 표준정규 | 전문가 | .8388 (.0415) | .8741 (.0393) | .8761 (.0402) |
| | | | 비전문가 | .7189 (.0998) | .7798 (.0961) | .7877 (.0998) |
| | | 지수 | 전문가 | .7727 (.0494) | .8175 (.0448) | .8185 (.0471) |
| | 비전문가 | | .6639 (.0945) | .7323 (.0904) | .7334 (.0961) | |
| | 5 | 표준정규 | 전문가 | .8920 (.0322) | .9025 (.0307) | .9048 (.0312) |
| | | | 비전문가 | .8066 (.0925) | .8250 (.0800) | .8348 (.0842) |
| 지수 | | 전문가 | .8363 (.0421) | .8518 (.0387) | .8537 (.0411) | |
| | 비전문가 | .7529 (.0867) | .7780 (.0769) | .7803 (.0848) | | |
| 반분면접 | 3 | 표준정규 | 전문가 | .8154 (.0444) | .8534 (.0350) | .8580 (.0351) |
| | | | 비전문가 | .7031 (.0762) | .7647 (.0691) | .7744 (.0715) |
| | | 지수 | 전문가 | .7403 (.0543) | .8008 (.0381) | .8003 (.0407) |
| | 비전문가 | | .6363 (.0773) | .7125 (.0712) | .7110 (.0762) | |
| | 5 | 표준정규 | 전문가 | .8699 (.0352) | .8760 (.0326) | .8818 (.0328) |
| | | | 비전문가 | .7851 (.0683) | .8023 (.0590) | .8138 (.0609) |
| 지수 | | 전문가 | .8089 (.0481) | .8315 (.0370) | .8309 (.0398) | |
| | 비전문가 | .7184 (.0746) | .7525 (.0650) | .7516 (.0709) | | |

표 2.2 실제 점수와 세 가지 표준화 방법들 간의 순위상관계수 (상관계수=0.4)

| 면접유형 | 면접자수 | 분포 | 면접자집단 | 표준화 방법 | | |
|------|------|---------------|---------------|---------------|---------------|---------------|
| | | | | 점사평균 | 순위평균 | z-점수평균 |
| 전체면접 | 3 | 표준정규 | 전문가 | .8448 (.0384) | .8779 (.0369) | .8797 (.0376) |
| | | | 비전문가 | .7306 (.1135) | .7827 (.1043) | .7895 (.1073) |
| | | 지수 | 전문가 | .7913 (.0580) | .8312 (.0522) | .8346 (.0529) |
| | 비전문가 | | .6727 (.1353) | .7319 (.1265) | .7282 (.1334) | |
| | 5 | 표준정규 | 전문가 | .8938 (.0311) | .9026 (.0304) | .9047 (.0310) |
| | | | 비전문가 | .8048 (.0976) | .8132 (.0936) | .8209 (.0972) |
| 지수 | | 전문가 | .8563 (.0392) | .8664 (.0395) | .8702 (.0400) | |
| | 비전문가 | .7694 (.1076) | .7840 (.1043) | .7800 (.1122) | | |
| 반분면접 | 3 | 표준정규 | 전문가 | .8184 (.0461) | .8559 (.0318) | .8602 (.0325) |
| | | | 비전문가 | .7016 (.0879) | .7537 (.0840) | .7623 (.0863) |
| | | 지수 | 전문가 | .7607 (.0584) | .8128 (.0428) | .8145 (.0434) |
| | 비전문가 | | .6519 (.1013) | .7206 (.0935) | .7139 (.0997) | |
| | 5 | 표준정규 | 전문가 | .8712 (.0356) | .8787 (.0298) | .8836 (.0306) |
| | | | 비전문가 | .7909 (.0710) | .7982 (.0673) | .8079 (.0695) |
| 지수 | | 전문가 | .8232 (.0468) | .8430 (.0353) | .8441 (.0365) | |
| | 비전문가 | .7396 (.0851) | .7633 (.0799) | .7571 (.0870) | | |

표 2.3 실제 점수와 세 가지 표준화 방법들 간의 순위상관계수 (상관계수=0.8)

| 면접유형 | 면접자수 | 분포 | 면접자집단 | 표준화 방법 | | |
|------|------|---------------|---------------|---------------|---------------|---------------|
| | | | | 점사평균 | 순위평균 | z-점수평균 |
| 전체면접 | 3 | 표준정규 | 전문가 | .8623 (.0340) | .8899 (.0301) | .8908 (.0301) |
| | | | 비전문가 | .7550 (.1466) | .7645 (.1400) | .7663 (.1409) |
| | | 지수 | 전문가 | .8259 (.0437) | .8592 (.0380) | .8626 (.0369) |
| | 비전문가 | | .7119 (.1712) | .7239 (.1604) | .7210 (.1698) | |
| | 5 | 표준정규 | 전문가 | .9064 (.0238) | .9115 (.0253) | .9126 (.0255) |
| | | | 비전문가 | .8368 (.1073) | .8008 (.1218) | .8033 (.1239) |
| 지수 | | 전문가 | .8851 (.0264) | .8888 (.0281) | .8925 (.0269) | |
| | 비전문가 | .8201 (.1142) | .7807 (.1265) | .7788 (.1364) | | |
| 반분면접 | 3 | 표준정규 | 전문가 | .8284 (.0455) | .8656 (.0293) | .8698 (.0298) |
| | | | 비전문가 | .7212 (.1092) | .7441 (.1029) | .7482 (.1044) |
| | | 지수 | 전문가 | .7874 (.0552) | .8369 (.0333) | .8397 (.0332) |
| | 비전문가 | | .6818 (.1290) | .7129 (.1148) | .7079 (.1241) | |
| | 5 | 표준정규 | 전문가 | .8794 (.0346) | .8845 (.0293) | .8883 (.0306) |
| | | | 비전문가 | .8134 (.0741) | .7845 (.0793) | .7892 (.0813) |
| 지수 | | 전문가 | .8514 (.0429) | .8645 (.0286) | .8664 (.0293) | |
| | 비전문가 | .7895 (.0823) | .7662 (.0856) | .7624 (.0958) | | |

이 표를 통해 우리가 상식적으로 알고 있는 사실을 확인할 수 있다. 그 사실은 전체면접이 반분면접보다, 5인 면접자가 3인 면접자보다, 정규분포가 지수분포보다, 전문가집단이 비전문가집단보다, 그리고 상관계수가 클수록 순위상관계수 값이 크다는 것이다.

세 방법 간의 순위상관계수 값을 비교해보면 그 차이가 그리 커 보이지는 않지만 다음과 같은 패턴을 확인할 수 있다. 정규분포에서는 z-점수평균법이 가장 좋고 그 다음으로 순위평균법, 절사평균법 순서인 것을 알 수 있다. 예외가 되는 것은 상관계수가 0.8인 비전문가집단의 5인 면접자인 경우로 절사평균법의 성능이 가장 좋고 다음으로 z-점수평균법, 순위평균법 순서가 된다. 따라서 z-점수평균법이 정규분포에서 좋은 성능을 발휘하는 것을 확인할 수 있었다. 지수분포에서는 전문가집단의 경우 z-점수평균법, 순위평균법, 절사평균법 순서이며, 비전문가집단의 경우 순위평균법, z-점수평균법, 절사평균법 순서인데 예외적으로 상관계수가 0.8인 5인 면접자에서는 절사평균법, 순위평균법, z-점수평균법 순서가 된다.

요약하면 z-점수평균법은 정규분포이거나 전문가집단인 경우에서 가장 좋은 성능을 보였다. 순위평균법은 정규분포에서 벗어난 지수분포인 경우 비전문가집단에서 가장 좋은 성능을 보였으며 절사평균법은 상관계수가 큰 0.8이면서 비전문가집단의 5인 면접자에서 가장 좋은 성능을 보였다.

순위평균법이 정규분포에서 많이 벗어나는 경우 좋은 성능을 발휘하리라고 기대했었는데 면접자가 전문가가 아닌 비전문가인 경우에서만 그렇다는 것을 확인할 수 있었다. 절사평균법은 상관계수가 0.8인 경우 면접자의 시각이 정확하지 않더라도 참값에 가까운 관측 면접 점수를 얻을 수 있어 간단하게 최저값과 최대값을 제외한 중간 3인의 평균을 사용하는 것이 다른 복잡한 변환과정을 거쳐 평균값을 사용하는 것보다 더 좋은 성능을 보이는 것으로 추측된다.

앞의 결론이 아주 작은 순위상관계수 값의 차이에 의한 결론이기 때문에, 통계적으로 의미가 있는 결론인지 알아보기 위해서 세 가지 방법을 두 개씩 짝을 지어 쌍체 t-검정 (paired t-test)을 수행하였다. 쌍체 t-검정의 결과를 정리한 것이 표 2.4에서 표 2.6에 나타나 있다. 독립표본 t-검정 대신에 쌍체 t-검정을 사용한 이유는 세 가지 방법의 값이 동일한 목표를 가지고 있기 때문에 상당히 큰 양의 상관을 가지기 때문이다.

표 2.4 세 가지 표준화 방법들 상호간의 쌍체 t-검정 (상관계수=0.0)

| 면접유형 | 면접자수 | 분포 | 면접자집단 | 표준화 방법 | | |
|------|------|------|-------|------------|------------|------------|
| | | | | z-점수 - 순위 | z-점수 - 절사 | 순위 - 절사 |
| 전체면접 | 3 | 표준정규 | 전문가 | 12.3706 ** | 44.0348 ** | 43.3433 ** |
| | | | 비전문가 | 21.9647 ** | 34.3091 ** | 32.9079 ** |
| | | | 지수 | 3.8530 ** | 45.427 ** | 46.7509 ** |
| | | 비전문가 | 전문가 | 1.8855 | 39.6395 ** | 42.5084 ** |
| | | | 비전문가 | 16.4989 ** | 30.8696 ** | 25.4466 ** |
| | | | 지수 | 26.943 ** | 31.458 ** | 19.9937 ** |
| | 5 | 표준정규 | 전문가 | 6.9537 ** | 34.2478 ** | 30.2226 ** |
| | | | 비전문가 | 3.5670 ** | 30.1229 ** | 25.8847 ** |
| | | | 지수 | 16.1698 ** | 31.035 ** | 27.5411 ** |
| | | 비전문가 | 전문가 | 26.3083 ** | 44.033 ** | 40.073 ** |
| | | | 비전문가 | -1.2006 | 37.0589 ** | 38.7832 ** |
| | | | 지수 | -2.4132 * | 40.931 ** | 44.6418 ** |
| 반분면접 | 3 | 표준정규 | 전문가 | 18.6921 ** | 10.3889 ** | 5.2622 ** |
| | | | 비전문가 | 31.7427 ** | 25.7337 ** | 15.0779 ** |
| | | | 지수 | -1.535 | 15.7677 ** | 16.7714 ** |
| | 5 | 표준정규 | 전문가 | -1.6578 | 24.9941 ** | 25.5392 ** |
| | | | 비전문가 | 18.6921 ** | 10.3889 ** | 5.2622 ** |
| | | | 지수 | 31.7427 ** | 25.7337 ** | 15.0779 ** |

* : 유의수준 0.05에서 유의한 차이를 보임, ** : 유의수준 0.01에서 유의한 차이를 보임

표 2.5 세 가지 표준화 방법들 상호간의 쌍체 t-검정 (상관계수=0.4)

| 면접유형 | 면접자수 | 분포 | 면접자집단 | 표준화 방법 | | |
|------|------|------|-------|-------------|------------|------------|
| | | | | z-점수 - 순위 | z-점수 - 결사 | 순위 - 결사 |
| 전체면접 | 3 | 표준정규 | 전문가 | 12.1119 ** | 44.9549 ** | 43.4552 |
| | | | 비전문가 | 19.215 ** | 30.345 ** | 28.8265 ** |
| | | 지수 | 전문가 | 12.7398 ** | 42.2199 ** | 40.6268 ** |
| | 5 | 표준정규 | 비전문가 | -6.3157 ** | 30.4618 ** | 34.5281 ** |
| | | | 전문가 | 14.7463 ** | 25.6124 ** | 21.2706 ** |
| | | 지수 | 비전문가 | 22.9011 ** | 12.2685 ** | 6.7507 ** |
| 반분면접 | 3 | 표준정규 | 전문가 | 15.3773 ** | 25.7275 ** | 17.67 ** |
| | | | 비전문가 | -6.4426 ** | 7.8808 ** | 11.9334 ** |
| | | 지수 | 전문가 | 15.1247 ** | 29.6923 ** | 26.6073 ** |
| | 5 | 표준정규 | 비전문가 | 22.6293 ** | 37.1475 ** | 33.3084 ** |
| | | | 전문가 | 4.2891 ** | 33.5755 ** | 33.0865 ** |
| | | 지수 | 비전문가 | -11.0799 ** | 34.9 ** | 40.428 ** |
| 전체면접 | 3 | 표준정규 | 전문가 | 16.5054 ** | 10.8649 ** | 6.6336 ** |
| | | | 비전문가 | 28.1939 ** | 13.5608 ** | 5.9242 ** |
| | | 지수 | 전문가 | 2.7817 ** | 15.3516 ** | 14.7183 ** |
| | 5 | 표준정규 | 비전문가 | -10.0212 ** | 11.8164 ** | 17.12 ** |
| | | | 전문가 | 16.5054 ** | 10.8649 ** | 6.6336 ** |
| | | 지수 | 비전문가 | 28.1939 ** | 13.5608 ** | 5.9242 ** |

* : 유의수준 0.05에서 유의한 차이를 보임, ** : 유의수준 0.01에서 유의한 차이를 보임

표 2.6 세 가지 표준화 방법들 상호간의 쌍체 t-검정 (상관계수=0.8)

| 면접유형 | 면접자수 | 분포 | 면접자집단 | 표준화 방법 | | |
|------|------|------|-------|------------|-------------|-------------|
| | | | | z-점수 - 순위 | z-점수 - 결사 | 순위 - 결사 |
| 전체면접 | 3 | 표준정규 | 전문가 | 7.2329 ** | 47.1531 ** | 45.6167 ** |
| | | | 비전문가 | 6.3598 ** | 6.299 ** | 5.4188 ** |
| | | 지수 | 전문가 | 18.2001 ** | 47.3766 ** | 43.1446 |
| | 5 | 표준정규 | 비전문가 | -4.4046 ** | 4.6875 ** | 6.8055 ** |
| | | | 전문가 | 10.011 ** | 16.6806 ** | 13.4757 ** |
| | | 지수 | 비전문가 | 9.3035 ** | -16.9575 ** | -19.0794 ** |
| 반분면접 | 3 | 표준정규 | 전문가 | 22.982 ** | 19.8567 ** | 8.772 |
| | | | 비전문가 | -3.1801 ** | -20.8506 ** | -22.6457 |
| | | 지수 | 전문가 | 14.2955 ** | 30.4363 ** | 27.0224 |
| | 5 | 표준정규 | 비전문가 | 12.1559 ** | 15.9217 ** | 13.5696 |
| | | | 전문가 | 8.1497 ** | 32.3339 ** | 30.7742 |
| | | 지수 | 비전문가 | -7.7902 ** | 13.3932 ** | 16.5598 ** |
| 전체면접 | 3 | 표준정규 | 전문가 | 12.5441 ** | 7.4883 ** | 4.2987 ** |
| | | | 비전문가 | 14.1311 ** | -14.2979 ** | -17.4413 ** |
| | | 지수 | 전문가 | 5.4326 ** | 10.7708 ** | 9.4469 |
| | 5 | 표준정규 | 비전문가 | -6.0283 ** | -14.7429 ** | -13.7683 ** |
| | | | 전문가 | 5.4326 ** | 10.7708 ** | 9.4469 |
| | | 지수 | 비전문가 | -6.0283 ** | -14.7429 ** | -13.7683 ** |

* : 유의수준 0.05에서 유의한 차이를 보임, ** : 유의수준 0.01에서 유의한 차이를 보임

이 표에서는 유의수준 0.05 혹은 0.01에서 유의한 차이를 보이는 쌍체 t-검정에는 *와 **라고 표시하고 있다. 이 표를 살펴보면 아주 작은 차이로 보였던 순위상관계수 값들의 차이가 소수 몇 개의 차이를 제외하고는 대부분 유의수준 0.01에서 유의한 차이를 보이는 것으로 나타났다. 이러한 현상은 박철용 (2011)에서 관측되었듯이 세 가지 방법에 의한 최종 면접점수들 사이에 상관관계가 크고 모의실험 횟수가 1000으로 크기 때문에 기인한 것이라 판단된다.

3. 결론

이 연구에서는 면접 점수가 발생하는 통계적 모형을 사영에 근거하여 제시하였다. 피면접자의 개별 참값과 이와 관련된 변수값이 2차원 평면의 X와 Y축의 값으로 주어졌을 때, 심사위원의 시각을 X축과의 각도로 생각하여 이 축에 사영된 값으로서 심사위원의 면접 점수의 평균을 잡고 개인적 편향과 관측 오차를 더해져 심사위원의 관측 면접 점수가 얻어지게 된다. 이 통계적 모형을 사용하여 흔히 사용되고 있는 면접 점수 표준화 방법인 절사평균법, 순위평균법, z-점수평균법을 비교하였다. 이 모의실험에서는 두 가지 면접 형태, 두 가지 면접자수, 두 가지 면접자의 전문성 정도, 실제 점수와 관련된 변수 간의 분포 두 가지와 세 가지 상관계수가 고려되었다.

모의실험 결과 z-점수평균법은 정규분포이거나 전문가집단인 경우에서 가장 좋은 성능을 보였다. 순위평균법은 정규분포에서 벗어난 지수분포인 경우 비전문가집단에서 가장 좋은 성능을 보였으며 절사평균법은 상관계수가 큰 0.8이면서 비전문가집단의 5인 면접자에서 가장 좋은 성능을 보였다.

이 모의실험 설계에서 여러 가지를 변인을 시도하였지만 이 실험결과를 일반적으로 확대해석하는 것은 곤란하리라고 생각한다. 좀 더 일반적인 결론을 내리기 위해서는 모의실험 설계에서 각 변인 별 두 개 값의 변동이 기본인 이 모의실험보다 좀 더 많은 변동을 포함시켜야 하리라고 생각하기 때문이다.

이 연구의 추후 연구과제로 다음과 같은 것을 생각할 수 있을 것이라 생각한다. 먼저 면접 점수를 발생시키는 보다 정교한 모형이 필요할 수도 있다. 이 연구의 관측 면접 점수를 발생시키는 통계적 모형은 실제 점수가 0을 중심으로 분포하고 있으며 관측 면접 점수도 0을 중심으로 분포하는 것이다. 실제 점수의 중심점을 0이 아닌 다른 값으로 옮기면 관측 면접 점수의 순서에는 상관이 없지만 관측 면접 점수를 많이 왜곡시킬 가능성이 있다. 이 연구에서는 순위상관계수를 이용하여 표준화 방법 간의 비교를 하기 때문에 중심점 이동에 상관없는 결과를 얻지만 표준화 방법 간 비교를 위해 사용하는 통계량이 이러한 중심점 이동에 민감할 경우 이를 반영하는 보다 정교한 통계적 모형이 필요하리라 생각된다. 이와 연관된 문제로 실제 입시에서는 면접 점수를 표준화시킬 때 어떤 범위에 속하게 되는 것을 요구하는 경우가 많다. 예를 들어 면접점수를 100점 만점으로 정하여 최하 0점부터 최대 100점까지 취하게 되도록 요구하는 경우가 있을 것이다. 순위상관계수를 사용하는 경우에는 어떠한 단순증가함수를 취하더라도 동일한 모의실험 결과를 얻게 된다. 그러나 원점수에 기반한 상관계수는 사용하는 단순증가함수 형태에 따라 다른 값을 취하기 때문에 어떠한 함수를 취하는가 하는 것이 중요한 문제가 된다. 따라서 어떠한 함수를 사용하는 것이 합리적인지 연구하는 것도 필요하리라 생각된다.

참고문헌

- 민대기, 정지현 (2012). 원점수 순위와 표준점수 순위 비교를 통한 표준화 방법비교. <한국데이터정보과학회지>, **12**, 113-120.
- 박성현, 김준원, 박준오 (2000). 대학입시에서의 선택과목 점수 표준화에 관한 연구. <품질경영학회지>, **28**, 124-132.
- 박철용 (2011). 면접점수 표준화 방법 모의실험 비교. <한국데이터정보과학회지>, **22**, 189-196.
- 이재만, 차영준 (2012). 학생에 의한 강의평가에서 집단별 표준화변환을 이용한 강좌평가모형. <한국데이터정보과학회지>, **23**, 143-150.
- 허명희 (1994). 새 대학입시의 통계적 계획과 분석 - 문항분석과 선택과목 등화(표준점수제)를 중심으로. <한국통계학회논문집>, **1**, 215-225.
- 황형태 (2005). 대학수학능력시험에서 표준점수제의 개선방안에 대한 연구. <응용통계연구>, **18**, 521-532.
- 황형태, 이강섭, 이장택 (2004). 대학입시에서의 면접점수 표준화에 관한 연구. <대한수학교육학회지 시리즈 A: 수학교육>, **43**, 309-314.

- Johnson, R. A. and Wichern, D. W. (1992). *Applied multivariate statistical analysis*, 3rd edition, Prentice Hall, New Jersey.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81-93.

A statistical model for interview score based on projection

Cheolyong Park¹ · Hyun Wook Kim²

¹²Department of Statistics, Keimyung University

Received 22 April 2012, revised 5 May 2012, accepted 15 May 2012

Abstract

In this study, we propose a statistical model based on projection that generates interview scores. In this model, each interviewee's true score and its related variable are viewed as X, Y values respectively in the two dimensional plane, and each interviewer's score is the projected score of true score X to the axis rotated by some angle, which reflects the interviewer's perspective. Each interviewer's observed interview score is obtained by adding personal bias and observed error to this projected score. We compared commonly used standardizing methods of interview scores such as trimmed mean method, rank method, and z-score method based on the proposed statistical model. In this simulation, two types of interview methods, two numbers of interviewers, two degrees of interviewers' expertise and two distributions and three correlations between actual score and its related variable are all considered.

Keywords: Projection, rank correlation, rank mean, trimmed mean, z-score mean.

¹ Corresponding author: Professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea. E-mail: cypark1@kmu.ac.kr

² Master, Department of Statistics, Keimyung University, Daegu 704-701, Korea.