

외국어시험에 대한 통계적분석방법

김 일 광

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《교육방법을 혁신하는데 맞게 학교들에서 시험방법, 학생들의 실력평가방법을 개선하여야 합니다.》(《김정일선집》 증보판 제23권 399페이지)

현시기 외국어교육에서 외국어시험을 통하여 학생들의 외국어실력을 정확히 평가하자면 시험에 대한 통계적분석을 진행한데 기초하여 시험방법, 실력평가방법을 부단히 개선해나가야 한다.

외국어시험에 대한 통계적분석은 빈도수표, 평균점수와 최빈점수, 중위점수, 점수범위와 표준편차, Z점수와 상관결수, 난도와 구분도, 믿음도 등을 리용하여 진행할수 있다.

무엇보다먼저 학생들의 시험결과에 대한 초보적인 리해를 가지려면 빈도수표를 작성하여 리용하는것이 좋다. 빈도수표는 똑같은 점수를 맞은 학생들의 수를 계산하는 방식으로 작성한다. 실례로 50명의 학생들이 응시한 외국어시험의 결과를 그대로 라렬하여놓으면 그들의 시험결과를 한눈에 명확히 알아볼수 없다. 이런 경우에 빈도수표를 리용하여 시험점수들을 직관적으로, 리해하기 쉽게 정리할수 있다.

표 1	
점수	빈도수 (명)
1.0	4
1.5	4
2.0	5
2.5	4
3.0	9
3.5	5
4.0	6
4.5	7
5	6
50	

시험실천에서는 표 1의 실례에서처럼 간단하게 빈도수표를 작성할수 있게 시험점수자료가 주어지지 않는 경우가 더 많이 존재한다. 그러나 5점제를 리용하여 시험점수를 소수점아래 한자리까지 계산하면 가능한 모든 점수는 51가지이다. 그리고 한번의 외국어시험에 응시하는 학생수가 많을수록 시험결과자료는 더욱 복잡해지며 그 량도 많아지게 된다. 이런 경우에는 시험점수를 일정한 차이를 두고있는 여러개의 구간으로 나누고 매 구간의 점수빈도수를 나타내는 방식으로 빈도수표를 작성하는것이 보다 편리하다. 아래에 200명의 학생들의 외국어시험점수를 점수구간에 따라 작성한 빈도수표의 실례를 보여주었다.

표 2

점수구간	빈도수 (명)
0.00-0.49	0
0.50-0.99	2
1.00-1.49	4
1.50-1.99	13
2.00-2.49	23
2.50-2.99	35
3.00-3.49	32
3.50-3.99	27
4.00-4.49	26
4.50-5	38
200	

빈도수표는 외국어시험결과에 대한 점수자료들을 한눈에 알아볼수 있게 정리하여 여러가지 목적에 신속하면서도 정확히 리용할수 있게 해준다.

다음으로 학생들의 시험점수들을 가지고 통계적분석에 필요한 여러가지 수값들을 계산하는것이 중요하다. 그런 수값들에는 평균점수와 최빈점수, 중위점수가 있다.

평균점수는 하나의 외국어시험에 응시한 전체 학생들의 점수를 산수평균하여 계산한 값이다. 평균점수는 한 시험에 응시한 모든 학생들의 시험점수들을 가지고 계산한 값인것으로 하여 계산방법이 편리하고 쉽게 리해할수 있는 우점을 가지고있다. 그리고 한 학생집단의 시험결과를 대표할수 있는것으로 하여 시험의 난도나 학생들의 전반적수준에 대한 초보적인 표상을 가지는데 적지 않은 도움을 준다.

최빈점수는 제일 많은 학생들이 받은 점수를 말한다. 최빈점수는 계산이 쉽고 시험결과자료에서 지나치게 높거나 지나치게 낮은 특이값들의 영향을 적게 받는다. 그러므로 특이값의 영향을 적게 받는 최빈점수는 한 학생집단의 시험결과를 비교적 정확히 대표할수 있는 값으로 된다.

중위점수는 모든 시험점수들을 크기순서로 배열하였을 때 제일 가운데 놓이는 점수이다. 중위점수 역시 특이값의 영향을 받지 않으므로 한 학생집단의 시험결과를 대표하는 값으로 쓸수 있다.

평균점수와 최빈점수, 중위점수들을 가지고서는 각이한 점수들로 이루어진 시험결과자료에 대한 구체적이며 정확한 표상을 얻기 힘들다. 간단한 실험으로 평균점수와 최빈점수가 똑같은 두조의 학생집단의 시험결과자료를 대비하여 보기로 하자.

표 3

학생		1	2	3	4	5	6	7	8	9	10	평균점수
점수	1조	1.2	3.6	4.0	3.9	3.9	4.9	2.7	3.9	2.1	3.6	3.38
	2조	3.9	3.4	3.0	3.5	2.9	3.9	3.1	3.3	2.8	4.0	3.38

표 3을 보면 두 집단의 평균점수와 최빈점수는 다같이 3.38과 3.9이다. 그러나 시험 점수들의 분포상태는 상당히 차이난다. 이런 경우 두 조의 점수들을 평균점수나 최빈점수 하나만 가지고 비교한다면 시험결과를 정확히 분석할수 없다.

이런데로부터 점수범위와 표준편차를 계산하여 평균점수를 중심으로 하는 각 점수들의 분포상태를 확인하여야 할 필요성이 제기된다.

점수범위는 한 학생집단의 시험결과자료에서 제일 높은 점수에서 제일 낮은 점수를 뺀 값이다. 표 3의 시험결과자료에 기초하여 두 집단의 점수범위를 계산하면 1조는 3.7이고 2조는 1.2이다. 보는바와 같이 비록 평균점수나 최빈점수가 같은 두 집단이라고 하여도 점수범위는 무려 2.5의 차이가 있다. 이로부터 두 집단의 실력상태가 같지 않다는것을 쉽게 확인할수 있다. 그러나 점수범위는 개개의 점수들이 평균점수로부터 얼마만큼 떨어져있는가를 반영하지 못한다. 점수범위가 가지고있는 이러한 약점을 보충해주는 수값이 바로 표준편차이다.

표준편차는 평균점수와와의 관계에서 시험결과자료의 분포정도를 나타내는 값으로서 평균점수를 중심으로 한 개별적인 시험점수들의 집중과 분산상태를 반영한다. 시험점수를 x , 평균점수를 M , 학생수를 n 이라고 하면 표준편차의 계산공식은 다음과 같다.

$$SD = \sqrt{\frac{\sum(x-M)^2}{n}}$$

표준편차가 작을수록 시험점수들은 평균점수와 가까이 분포되어있으며 표준편차가 클수록 시험점수들은 평균점수로부터 넓게 분포되어있다. 표준편차는 학생집단의 시험결과를 분석하는데뿐만아니라 개별적인 학생이 받은 여러개의 시험점수들이나 개별적인 채점자의 여러차례의 채점결과를 분석하여 시험점수의 분포상태를 정확히 확인할수 있으며 따라서 채점의 믿음성을 평가하는데도 리용할수 있다.

표준편차의 계산공식은 어지간히 복잡하고 계산공정이 까다롭지만 컴퓨터의 표계산 프로그램에 들어있는 《STDEV.P》함수를 리용하면 표준편차를 쉽게 그리고 정확히 구할수 있다.

시험결과를 일정한 채점기준에 근거하여 평가된 점수는 시험결과의 발표와 학생에 대한 평가를 내리는데서 중요한 의의를 가지지만 외국어시험의 통계적분석에서는 일련의 문제점을 가진다.

첫째로, 하나의 시험을 이루는 부분적인 시험들의 채점결과들을 서로 정확히 비교할수 없다. 실례로 어떤 학생의 듣기시험점수가 3.8이고 말하기시험점수가 4.1인 경우 두개의 시험점수를 비교하여 어느 실력이 보다 높다는것을 정확히 판단할수 없다.

둘째로, 서로 다른 시기에 친 시험들의 결과를 정확히 비교할수 없다. 실례로 어떤 학생의 1학기 외국어시험점수가 4.2이고 2학기의 외국어시험점수가 4.5일 때 단순히 시험점수의 증가를 놓고 그의 외국어실력이 향상되었다고 판단할수 없다.

셋째로, 서로 다른 시기에 친 서로 다른 외국어지식이나 기능에 대한 시험들의 결과를 서로 종합하여 분석할수 없다.

우와 같은 문제점들이 존재하는 기본원인은 외국어시험들마다 목적과 기준이 서로 다르고 난도에서도 서로 차이난기때문이다. 이러한 문제점들을 극복하기 위하여 Z점수를 리용한다.

Z점수는 평균점수를 0으로, 표준편차를 1로 보는 점수를 말한다. Z점수는 시험점수에서 평균점수를 뺀 값을 표준편차로 나누는 방법으로 계산한다.

어느 한 외국어시험에 응시한 한 학생이 듣기시험에서 3.8을, 말하기시험에서 4.0의 점수를 맞았다고 하자. 이때 듣기시험에 응시한 학생들의 평균점수는 3.4, 표준편차는 0.36이고 말하기시험에 응시한 학생들의 평균점수는 4.2이고 표준편차는 0.51이다.

두 시험은 난도도 다르고 평균점수도 다른것으로 하여 이 학생의 시험점수만 가지고서는 어느 실력이 상대적으로 높은지 정확히 평가할수 없다. 따라서 Z점수를 리용하여야 정확히 비교할수 있다.

$$\text{듣기시험: } Z = \frac{3.8-3.4}{0.36} \approx 1.111 \quad \text{말하기시험: } Z = \frac{4.0-4.2}{0.51} \approx -0.392$$

계산결과에 의하여 이 학생의 듣기시험점수는 평균점수보다 표준편차 1.111만큼 높고 말하기시험점수는 평균점수보다 0.392만큼 낮다는것을 알수 있다. 이로부터 그의 듣기실력이 말하기실력보다 높다는 결론을 내릴수 있다.

외국어시험실천에서는 하나의 학생집단을 여러조로 나누고 하나의 목적과 같은 난도를 가진 시험들을 여러개의 안으로 갈라 실행하는 경우가 있다. 이런 경우에는 여러안으로 갈라진 개개의 시험들이 학생들의 외국어실력을 측정하는데서 동등하게 작용할수 있는가 하는것을 정확히 확인하여야 할 필요성이 제기된다. 이로부터 여러 시험들의 상관성을 과학적으로 분석할수 있도록 상관결수를 계산하여 리용하여야 한다. 상관결수를 계산하는 방법은 여러가지가 있지만 외국어시험실천에 적합한 계산방법은 두가지가 있다.

우선 등급상관분석법이 있다.

등급상관분석법은 학생들의 시험점수에 기초하여 결정한 순위를 가지고 상관결수를 계산하는 방법이다. 이때 다음과 같은 계산공식을 리용한다.

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

위의 계산공식에서 ρ 는 등급상관결수, d 는 순위차, n 은 학생수를 가리킨다.

등급상관결수를 계산할 때 같은 순위에 놓이는 학생들은 그 순위의 평균값을 순위값으로 정해주어야 한다. 실례로 3등에 2명의 학생이 놓이게 되면 그들의 순위는 3등과 4등에 해당하므로 순위값을 3.5로 계산한다.

또한 적차상관분석법이 있다.

적차상관분석법은 학생들의 순위가 아니라 시험점수를 직접 리용하여 상관결수를 구하는 방법이다. 적차상관분석법에서 리용하는 상관결수계산공식은 다음과 같다.

$$r = \frac{\Sigma(x_1 - M_1)(x_2 - M_2)}{nSD_1SD_2}$$

위의 공식에서 r 는 적차상관결수, x_1 과 x_2 는 시험1과 시험2의 점수, M_1 과 M_2 는 시험1과 시험2의 평균점수, SD_1 과 SD_2 는 시험1과 시험2의 표준편차를 가리킨다.

등급상관결수를 계산하는 공정과 비해볼 때 적차상관결수를 계산하는 공정은 상대적으로 복잡한것으로 하여 수동적으로 계산하는것보다 콤퓨터의 표계산프로그램을 리용하는 것이 편리하고 정확성도 보장할수 있다. 표계산프로그램을 가지고 계산할 때에는 적차

상관결수계산함수인 《PEARSON》함수를 리용한다.

상관결수는 -1과 +1사이에 놓이는 값으로서 +1에 가까울수록 상관성이 높은 시험으로 된다.

다음으로 외국어시험에서는 시험문제의 난도와 구분도를 정확히 계산하여야 한다.

난도는 시험문제의 어려운 정도를 가리키는 개념이다. 오직 0점과 1점으로 채점할수 있는 객관적시험문제들의 난도는 그 문제에 대하여 0점을 맞은 학생이 학생집단에서 차지하는 비율로 계산한다. 이때 0점을 맞은 학생수를 전체 학생수로 나누어 0과 1사이의 값범위에 놓이는 소수를 얻는다. 실례로 어떤 시험문제를 300명의 학생들이가운데서 150명이 정확히 풀지 못했다면 그 문제의 난도는 0.5이며 정확히 푼 학생이 6명밖에 없다면 그 문제의 난도는 0.98로 된다.

말하기와 글쓰기, 번역시험문제와 같은 주관적시험문제들의 난도는 그 문제에 할당된 최고점수와 학생수를 곱한 값에서 모든 학생들이 그 문제의 풀이결과에 대하여 받은 점수들의 총합을 덜고 그 뺀 값을 학생수로 나눈 다음 다시 최고점수로 나누는 방법으로 계산한다. 실례로 최고점수가 5점인 주관적시험문제에 대한 채점점수들의 총합이 375점이고 응시한 학생수가 100명이라면 이 문제의 난도는 $(500-375) \div 100 \div 5 = 0.25$ 로 된다.

이렇게 시험문제의 난도를 정확한 수값으로 계산하면 시험의 목적에 정확히 부합되는 유효성과 믿음성이 높은 시험을 작성하는데 매우 유리하다. 실례로 어떤 외국어교육과정을 끝마치고 졸업생들가운데서 제일 우수한 10%의 학생들을 선발하기 위한 시험을 작성하는 경우 난도가 낮은 문제들은 배제하고 난도가 높은 문제들 즉 0.9이상의 난도를 가진 시험문제들로 시험을 작성하여 그 시험의 유효성과 믿음성을 정확히 보장할수 있다.

난도는 시험문제의 쉽거나 어려운 정도를 리해하는데서 큰 도움이 되지만 학생들의 외국어실력을 어느 정도로 정확히 구분하여줄수 있는가에 대하여서는 일반적인 표상밖에 안겨주지 못한다.

이로부터 실력이 높은 학생과 실력이 낮은 학생을 얼마나 잘 갈라주는가를 정량적으로 보여줄수 있는 지표인 구분도를 계산한다. 구분도가 높은 문제들로 작성된 시험은 믿음성도 높으므로 시험문제의 구분도를 정확히 분석하는것은 외국어시험실천에서 중요한 문제로 나선다.

구분도는 난도와 달리 0을 중심으로 정의 값이나 부의 값을 가질수 있다. 다시말하여 학생들의 실력을 전혀 구별해주지 못하는 문제 즉 실력이 높은 학생이나 실력이 낮은 학생이 다같이 정확히 풀수 있는 문제의 구분도는 0으로 되며 실력이 높은 학생이 정확히 풀수 있고 실력이 낮은 학생이 정확히 풀수 없는 문제의 구분도는 정의 값을 가지게 된다. 반대로 현실적인 가능성은 매우 낮지만 실력이 낮은 학생들이 잘 풀수 있지만 실력이 높은 학생들이 잘 풀수 없는 문제들이 있을수도 있다. 이런 경우에는 구분도가 부의 값을 가지게 된다. 구분도를 계산하는 방법은 다음과 같다.

첫째로, 시험점수에 따라 학생들의 순위를 결정하고 점수가 낮아지는 순서로 배열한다.

둘째로, 전체 학생들을 상, 중, 하의 3개 집단으로 구분한다. 이때 상과 하의 집단에 똑같은 수의 학생들이 들어가게 구분하여야 한다.

셋째로, 상의 집단에서 시험문제를 정확히 푼 학생수에서 하의 집단에서 시험문제를 정확히 푼 학생수를 뺀다.

넷째로, 위에서 나온 결과를 상의 집단(혹은 하의 집단)의 학생수로 나눈다.

시험의 믿음성을 높여 학생들의 외국어실력을 정확히 평가하자면 다음으로 외국어시험의 믿음성을 하나의 수값으로 나타낼수 있는 믿음도를 계산하여 리용하는것이 필요하다.

믿음도를 계산하려는 시험을 2개의 똑같은 부분으로 나눈다. 실례로 100개의 문제로 구성된 시험이라면 홀수번호문제들과 짝수번호문제들로 가를수 있다. 이렇게 만든 두개의 시험을 한 학생집단에 적합한 시간간격을 두고 차례로 실행한다. 두 시험의 채점결과가 나오면 다음과 같은 공정을 거쳐 전체 시험의 믿음도를 계산한다.

① 두 시험결과들사이의 등급상관결수(ρ)를 구한다.

② 스피어먼-브라운공식을 리용하여 등급상관결수값으로부터 믿음도를 계산한다. 스피어먼-브라운공식은 다음과 같다.

$$R = \frac{2\rho}{1+\rho}$$

하나의 외국어시험을 동등한 2개의 시험으로 가르고 한 학생집단에 서로 다른 기회에 실행할수 없는 경우에는 한번의 시험으로 얻어진 문제풀이행위를 두개의 시험처럼 갈라서 평가하여 학생 1인당 두개의 평가점수를 매긴다. 실례로 홀수번호문제들에 한하여 하나의 시험점수를 내고 짝수번호문제들에 한하여 다른 하나의 시험점수를 낸다. 이렇게 얻은 두조의 시험점수를 전체 시험을 두번 친것처럼 리용하여 믿음도를 계산한다. 이 계산방법이 실천에서 정확히 작용할수 있도록 하자면 시험문제들에 대한 구체적인 대비분석을 통하여 하나의 시험을 완전히 동등한 두 부분으로 똑같이 나누어야 한다.

시험의 믿음도를 계산할 때에는 다른 학생들보다 시험점수가 훨씬 높거나 혹은 훨씬 낮은 학생들이 있는 경우 믿음도가 잘못 계산되어나올수 있다는데 주의를 돌려야 한다. 그러므로 이런 경우에는 최대특이값과 최소특이값으로 되는 시험점수들을 계산에서 배제하여야 한다.

기준도달시험에서는 오직 두가지 평가결과 즉 합격과 불합격으로 시험결과가 주어지게 되므로 평가결과의 일관성을 놓고 믿음도를 계산하여야 한다.

① 두차례의 시험에서 다 합격으로 평가된 학생수를 확인한다.

② 두차례의 시험에서 다 불합격으로 평가된 학생수를 확인한다.

③ 위의 두 수를 더한 값을 전체 학생수로 나눈다.

실례로 50명의 학생이 응시한 두차례의 기준도달시험에서 두번 다 합격으로 평가된 학생이 18명이고 두번 다 불합격으로 평가된 학생이 15명이라면 이 시험의 믿음도는 아래와 같이 계산할수 있다.

$$R = (18 + 15) \div 50 = 0.66$$

이상에서 외국어시험에 대한 몇가지 통계적분석방법에 대하여 보았다.

우리는 외국어시험에 대한 통계적분석에 대한 학술적연구를 더욱 심화시키고 외국어 시험방법을 혁신함으로써 학생들을 유능한 혁명인재들로 튼튼히 준비시키는데 적극 이바지하여야 할것이다.